# Spatial Panel Data Model with Error Dependence: A Bayesian Separable Covariance Approach

Samantha Leorato[*] and Maura Mezzetti[†]

**Abstract.** A hierarchical Bayesian model for spatial panel data is proposed. The idea behind the proposed method is to analyze spatially dependent panel data by means of a separable covariance matrix. Let us indicate the observations as $y_{it}$, in $i = 1, \ldots, N$ regions and at $t = 1, \ldots, T$ times, and suppose the covariance matrix of $\mathbf{y}$, given a set of regressors, is written as a Kronecker product of a purely spatial and a purely temporal covariance. On the one hand, the structure of separable covariances dramatically reduces the number of parameters, while on the other hand, the lack of a structured pattern for spatial and temporal covariances permits capturing possible unknown dependencies (both in time and space). The use of the Bayesian approach allows one to overcome some of the difficulties of the classical (MLE or GMM based) approach. We present two illustrative examples: the estimation of cigarette price elasticity and of the determinants of the house price in 120 municipalities in the Province of Rome.

**Keywords:** Bayesian inference, Kronecker product, separable covariance matrix, inverse Wishart distribution, spatial–temporal dependence.

## 1 Introduction

In recent years, there has been a growing interest in studying panel data theory. The great advantage of panel data is that there is more than one realization of the observable processes, thus allowing for various forms of heterogeneity/dependence that are not possible for cross-section or time-series data separately. When for each time unit the observed variables refer to a spatial aggregate (region, country or municipality) they are usually referred to as spatial panel data. The development of spatial econometric models conveys particular emphasis on the role of either spatial structure (heterogeneity) or interactions in space (spatial autocorrelation). The fact that distance affects economic behavior is the main reason for a variable associated with a specific location being dependent on variables at other locations. In this perspective, different types of interaction effects may be considered: endogenous interaction effects, exogenous interaction effects or interaction effects among error terms. Usually the literature refers to spatial lag model for the former situation, and to spatial error model for the latter (see, for instance, Elhorst (2010)).

This paper complements the literature on models with error correlation for spatial panel data; in particular, we propose a hierarchical Bayesian model that accounts for

[*]Dep. Economics and Finance, Tor Vergata University, Rome, samantha.leorato@uniroma2.it
[†]Dep. Economics and Finance, Tor Vergata University, Rome, maura.mezzetti@uniroma2.it

both spatial and temporal error autocorrelations. To the authors' best knowledge, the first spatial panel data model allowing for both spatial and temporal dependences is due to Baltagi et al. (2007), who extended the spatial autoregressive error model by adding an autocorrelation structure on the errors at different time units. One of the first papers dealing with a Bayesian approach for the estimation of spatial panel data with autocorrelated errors is Zheng et al. (2008), who proposed a hierarchical Bayesian estimation of the model of Baltagi et al. (2007). Since then the spatial panel data literature has been growing exponentially fast. Most of the contributions, though, fall into a classical framework (mainly MLE or GMM based). In contrast, the Bayesian contributions to spatial panel models have been relatively few. One of the explicit aims of this work is therefore to fill the gap and help define a Bayesian alternative to the classical spatial–temporal error model.

As for the model of Zheng et al. (2008), our model considers both spatial and temporal heterogeneity, by the inclusion of fixed or random effects in a single model. However, unlike Zheng et al. (2008), we do not assign *a priori* any specific structure to the covariance matrix. The key assumption in our approach is the separability of the spatial and temporal components of the covariance (conditional on eventual heterogeneities of the covariates). The covariance matrix of the dependent variables thus takes the form of a Kronecker product of two covariance matrices (temporal and spatial). This assumption incorporates the main models considered in the spatial error dependence approach.

Several authors observed many advantages of the use of Kronecker product for the decomposition of the covariance for analyzing different types of data. For example, it has been applied to multivariate longitudinal data (see, for instance, Naik and Rao (2001) or Shults (2000)); some spatio-temporal models have also been taken into account in geological and biomedical applications (see Aston and Gunn (2005) or Kyriakidis and Journel (1999)). A well-known limit of separable covariance functions is that they do not allow for space–time interactions in the covariance (e.g., see Kyriakidis and Journel (1999) and Cressie and Huang (1999)). Nevertheless, statistical space–time modeling has been using separable covariance functions to a large extent. The main advantage is that the structure of a separable covariance drastically reduces the number of parameters in the covariance matrix, and therefore computational procedures for large space–time data sets are simplified. Furthermore, Genton (2007) studied the nearest Kronecker approximation of non-separable spatial–temporal covariance matrices, in the Frobenius norm.

A particular factorizable covariance matrix arises naturally in the model proposed by Baltagi et al. (2007), where the spatial and temporal covariances depend on scalar parameters measuring spatial and temporal dependencies. Therefore the hypotheses of Baltagi et al. (2007) (as well as Zheng et al. (2008)) can be derived by our model assumptions. The assumption of a structured covariance matrix, typically depending on a small fixed number of parameters, is more parsimonious and has advantages both in terms of computational burden and interpretability. On the other hand, according to this assumption, the pattern of the spatial dependence is given, usually governed by geographical proximity. It is however possible that some social, political or economical factors determine the proximity of two regions or countries, and these factors may not be

known at all. Therefore, our proposal of unstructured spatial and temporal covariances is expected to be more robust against misspecification of the error components.

The rest of the paper is structured as follows: after a brief review of the main competing models for spatial panel data in Section 1.1, in Section 2 we present our hierarchical model, and derive the posterior distributions in the general case (when we impose no structure on the spatial and temporal components). Section 3 considers the models obtained when the spatial or temporal patterns are specified. Two examples are treated as benchmarks: the autoregressive matrix for the temporal dimension and a spatial covariance based on the matrix exponential. After showing computational aspects in Section 4, Section 5 presents applications to simulated and real data, while Section 6 is devoted to sensitivity analysis. In particular, in Section 6.2 robustness to deviation from separability in covariance structure is checked. Finally, Section 7 concludes.

## 1.1   The existing models

In the last decade, many authors proposed alternative methods for the estimation of different specifications of spatial panel data models. Here we briefly review the main models for spatial panel data considered in the literature, with a particular focus on models with spatial autocorrelations. A deeper presentation of the different models can be found in Elhorst (2010), Anselin et al. (2008) and more recently in Lee and Yu (2012).

The alternative models can be divided into different classes, according to the nature of the spatial interaction effects considered, the presence or absence of dynamic effects, and the type of time- or space-specific effects. Concerning the spatial dimension, the model may contain spatially lagged dependent variables (spatial lag model), spatially lagged regressors (Durbin model) or spatial error autocorrelation (SEM). The spatial interaction in general is specified via a spatial autoregressive process that depends on a weight matrix $\mathbf{W}$. The weight matrix is given a priori and is typically either an adjacency matrix, that is, a matrix with zeros on the main diagonal and ones corresponding to neighboring couples of regions (usually normalized by rows), or a decreasing function of distances between regions.

Following Sarafidis and Wansbeek (2012), the main spatial error models are the spatial autoregressive (SAR), the spatial moving average (SMA) and the spatial error component (SEC) models. These spatial models can be estimated, for example, using a generalized method of moments (GMM) approach (see, e.g., Kapoor et al. (2007) and Kelejian and Prucha (2010)). In models with spatial and temporal covariances, the time dependence is usually represented by an autoregressive process as, for example, in Baltagi et al. (2003), Kapoor et al. (2007).

Spatial panel data models can include space specific effects, to account for heterogeneity in space due to unobserved characteristics of each spatial unit considered. Similar effects can be included in time that control spatially-invariant variables whose omission could bias the estimates in a time-series analysis. A model could include simultaneously space- and time-specific effects. All the different specifications described above can be

| Non-dynamic models | Dynamic models |
|---|---|
| Spatial lag model (SLM) | Spatial lag model (DSLM) |
| $\mathbf{y}_t = \boldsymbol{\mu}_t + \boldsymbol{\lambda}\mathbf{W}\mathbf{y}_t + \boldsymbol{\beta}'\boldsymbol{x}_t + \mathbf{u}_t$ | $\mathbf{y}_t = \boldsymbol{\mu}_t + \rho\mathbf{y}_{t-1} + \boldsymbol{\lambda}\mathbf{W}\mathbf{y}_t + \boldsymbol{\beta}'\boldsymbol{x}_t + \mathbf{u}_t$ |
| Spatial error models (SEM) | Spatial error models (DSEM) |
| $\mathbf{y}_t = \boldsymbol{\mu}_t + \boldsymbol{\beta}'\boldsymbol{x}_t + \boldsymbol{\varepsilon}_t$ | $\mathbf{y}_t = \boldsymbol{\mu}_t + \rho\mathbf{y}_{t-1} + \boldsymbol{\beta}'\boldsymbol{x}_t + \boldsymbol{\varepsilon}_t$ |
| (SMA) $\quad \boldsymbol{\varepsilon}_t = \lambda\mathbf{W}\mathbf{u}_t + \mathbf{u}_t$ | (SMA) $\quad \boldsymbol{\varepsilon}_t = \lambda\mathbf{W}\mathbf{u}_t + \mathbf{u}_t$ |
| (SAR) $\quad \boldsymbol{\varepsilon}_t = \lambda\mathbf{W}\boldsymbol{\varepsilon}_t + \mathbf{u}_t$ | (SAR) $\quad \boldsymbol{\varepsilon}_t = \lambda\mathbf{W}\boldsymbol{\varepsilon}_t + \mathbf{u}_t$ |
| (SAR+AR) $\quad \boldsymbol{\varepsilon}_t = \lambda\mathbf{W}\boldsymbol{\varepsilon}_t + \mathbf{v}_t$ | |
| $\mathbf{v}_t = \rho\mathbf{v}_{t-1} + \mathbf{u}_t$ | |
| (SEC) $\quad \boldsymbol{\varepsilon}_t = \lambda\mathbf{W}\boldsymbol{\xi}_t + \mathbf{u}_t$ | (SEC) $\quad \boldsymbol{\varepsilon}_t = \lambda\mathbf{W}\boldsymbol{\xi}_t + \mathbf{u}_t$ |
| Durbin model (DM) | Durbin model (DDM) |
| $\mathbf{y}_t = \boldsymbol{\mu}_t + \boldsymbol{\beta}'\boldsymbol{x}_t + \boldsymbol{\theta}\mathbf{W}\boldsymbol{x}_t + \mathbf{u}_t$ | $\mathbf{y}_t = \boldsymbol{\mu}_t + \rho\mathbf{y}_{t-1} + \boldsymbol{\beta}'\boldsymbol{x}_t + \boldsymbol{\theta}\mathbf{W}\boldsymbol{x}_t + \mathbf{u}_t$ |

Table 1: Specifications of the main spatial panel data models, where $\mathbf{u}_t$ is a vector of $N$ independent $N(0, \sigma^2)$ random variables and $\xi_{it}$ is a zero-mean random component uncorrelated with $u_{it}$ and $x_{it}$.

either non-dynamic or dynamic. A dynamic model consists in the inclusion of time-lagged values of the dependent variable. Table 1 presents a schematized view of all the models described above. Although far from being exhaustive, these models represent a good synthesis of the main models used in the spatial panel data literature. Standard spatial error models do not account for covariances in the time dimension. However, Table 1 also includes the model proposed by Baltagi et al. (2007) as an extension of the static SAR model. To the author's best knowledge, no dynamic counterpart has ever been proposed. As we stressed above, inference on all models in Table 1 mainly (if not exclusively) follows the non-Bayesian paradigm. In all the models below, the vector of specific effects $\boldsymbol{\mu}_t = (\mu_{1t}, \ldots, \mu_{Nt})'$ is equal to $\mu_{it} = \alpha_i$ for the spatial effects only, $\mu_{it} = \eta_t$, in the case of the temporal specific effect, and $\mu_{it} = \alpha_i + \eta_t$ when the model includes both. In all these cases, the effects can be assumed to be fixed or random. In the random effects model, $\alpha_i$ and $\eta_t$ are treated as random variables that are assumed to be mutually independent and independent with the errors $\varepsilon_{it}$.

## 2   The conditionally separable autocovariance model

Let $\mathbf{y} = \{y_{it}\}$ indicate the column vector $(N \times T)$, where $N$ is the number of spatial units and $T$ the number of time points and let $\boldsymbol{x} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_K)$, where $\boldsymbol{x}_k$ denotes the vector of $N \cdot T$ observations relative to the $k$th regressor (excluding time or space dummies). Moreover, let $\boldsymbol{\beta}$ be the vector of $K$ regression coefficients and $\boldsymbol{\mu} = (\boldsymbol{\mu}'_1, \ldots, \boldsymbol{\mu}'_T)'$ be an $N \cdot T$ column vector of fixed or random effects either regional $(\mu_{it} = \alpha_i)$, temporal $(\mu_{it} = \eta_t)$ or both $(\mu_{it} = \eta_t + \alpha_i)$.

In this paper, we propose a Bayesian approach for the estimation of a spatial panel data model, assuming a conditionally separable autocovariance matrix for the errors (CSA).

Specifically, the CSA model is a hierarchical linear model,

$$\mathbf{y} = \boldsymbol{x\beta} + \boldsymbol{\mu} + \boldsymbol{\varepsilon},$$
$$\mathbf{y} \mid \boldsymbol{x}, \boldsymbol{\beta}, \boldsymbol{\mu}, \sigma_\varepsilon^2, \boldsymbol{\Psi}, \boldsymbol{\Phi} \sim N(\boldsymbol{x\beta} + \boldsymbol{\mu}, \boldsymbol{\Sigma}), \quad (1)$$
$$\boldsymbol{\Sigma} = \mathrm{var}(\mathbf{y} \mid \boldsymbol{x}, \boldsymbol{\beta}, \boldsymbol{\mu}, \boldsymbol{\Psi}, \boldsymbol{\Phi}) = \boldsymbol{\Psi} \otimes \boldsymbol{\Phi},$$

namely the covariance matrix $\boldsymbol{\Sigma}$ is the Kronecker product of two symmetric positive definite (p.d.) matrices of sizes $T$ and $N$, respectively. In order to write the likelihood function, it is more convenient to rewrite the term $\boldsymbol{\mu}$ in (1) as

$$\boldsymbol{\mu} = (\mathbf{I}_T \otimes \boldsymbol{\imath}_N)\boldsymbol{\alpha} + (\boldsymbol{\imath}_T \otimes \mathbf{I}_N)\boldsymbol{\eta},$$

where $\boldsymbol{\alpha}$ and $\boldsymbol{\eta}$ are the vectors of space and time effects, respectively, while $\mathbf{I}_T \otimes \boldsymbol{\imath}_N$ and $\boldsymbol{\imath}_T \otimes \mathbf{I}_N$ are the corresponding matrices of dummies. Appropriately specifying $\boldsymbol{\Psi}$ and $\boldsymbol{\Phi}$ gives all the SEM models in Table 1. More precisely, the errors in (1) can be equivalently written as $\varepsilon_t = c_t \mathbf{D} \mathbf{u}_t$ and $\varepsilon_i = d_t \mathbf{C} \mathbf{u}_i$, where $\boldsymbol{\Psi} = \mathbf{C}\mathbf{C}^\top$ and $\boldsymbol{\Phi} = \mathbf{D}\mathbf{D}^\top$ and $c_t$ and $d_i$ are the $t$th and $i$th elements of the main diagonals of $\mathbf{C}$ and $\mathbf{D}$, respectively.

The introduction of a separable covariance matrix yields the intuitive interpretation of the two components $\boldsymbol{\Psi}$ and $\boldsymbol{\Phi}$ as the spatial and temporal covariance matrices, respectively. The covariances between the observations of the response variable are therefore the following:

$$cov(y_{it}, y_{js} | \boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\eta}, \boldsymbol{\Psi}, \boldsymbol{\Phi}) = \phi_{i,j}\psi_{t,s}. \quad (2)$$

Taking one of the two matrices as diagonal, we have the model with either the temporal or the spatial dependence considered separately. In particular, one obtains the well known SUR model (Zellner (1962)), when $\boldsymbol{\Phi}$ is the identity matrix. Model (1) in particular coincides with the one used in Zheng et al. (2008) if $\boldsymbol{\Psi}$ and $\boldsymbol{\Phi}$ are autoregressive and spatial autoregressive matrices, respectively.

Various authors (see, for instance, Mitchell et al. (2006)) noted that a lack of identifiability can result when covariance is defined through Kronecker product. The indeterminacy stems from the fact that given $\boldsymbol{\Sigma} = \boldsymbol{\Psi} \otimes \boldsymbol{\Phi}$, $\boldsymbol{\Psi}$ and $\boldsymbol{\Phi}$ are not unique since, for any $k > 0$, $k\boldsymbol{\Psi} \otimes (1/k)\boldsymbol{\Phi} = \boldsymbol{\Psi} \otimes \boldsymbol{\Phi}$. As a consequence, the likelihood in the Gaussian model (1) is invariant under scale transformations of $\boldsymbol{\Phi}$ and $\boldsymbol{\Psi}$. This non-identifiability in the frequentist contest is generally fixed by rescaling one of the factor-specific covariance matrices so that one of its diagonal nonzero elements is equal to 1, (see, for instance, Srivastava et al. (2008)), for example, by setting $\Psi_{11} = 1$ or $\Phi_{11} = 1$. Historically, Bayesian statisticians have been less interested in the concept since, strictly speaking, any parameter having a proper prior distribution is estimable, thus, by using a Bayesian hierarchical model with proper priors for $\boldsymbol{\Psi}$ and $\boldsymbol{\Phi}$, we can ignore this problem. In fact, as noted by Lindley (1972), the problem of non-identification is not a real difficulty in practice in the Bayesian context, as long as the posterior distribution is well defined, which in general occurs if the prior is a proper distribution. Moreover, the posterior distribution of coefficients $\beta$ depends on $\boldsymbol{\Sigma} = \boldsymbol{\Psi} \otimes \boldsymbol{\Phi}$, thus, is not affected by changes in the scale of $\boldsymbol{\Psi}$ and $\boldsymbol{\Phi}$.

The lack of information about the possible source of correlation both in the spatial and in the serial dimensions, especially in the unobservables, is the main reason for

considering the extension of the traditional SEM models to a more general assumption of separable spatial and temporal covariance matrices. Similar considerations also motivated Pesaran (2004) and Pesaran and Tosetti (2011) to propose an alternative residual multifactor approach, where spatial dependence is determined by a small number of unobserved common factors.

As an alternative to the main model illustrated in the following subsections, Section 3 presents the case of structured matrices $\boldsymbol{\Psi}$ and $\boldsymbol{\Phi}$. In this case, prior probabilities simplify to $p(\boldsymbol{\Psi}) = p(\rho)$ and $p(\boldsymbol{\Phi}) = p(\lambda)$ where $\rho$ and $\lambda$ are scalar parameters. We refer to the case when the structure of $\boldsymbol{\Psi}$ (or $\boldsymbol{\Phi}$) is known as *structured* and when it is unknown as *unstructured*.

In all the specifications considered below, we assume prior independence of the parameters of the model, namely that

$$p(\boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\eta}, \boldsymbol{\Phi}, \boldsymbol{\Psi}) = p(\boldsymbol{\Psi})p(\boldsymbol{\Phi})p(\boldsymbol{\beta})p(\boldsymbol{\alpha})p(\boldsymbol{\eta}). \tag{3}$$

Most of the prior distributions in (3) are chosen within the conjugate families. Therefore, all $\boldsymbol{\beta}$, $\boldsymbol{\eta}$ and $\boldsymbol{\alpha}$ are assumed to have a Gaussian prior so that

$$\boldsymbol{\beta} \mid \boldsymbol{\beta}_0, \mathbf{H} \sim \mathcal{N}_k (\boldsymbol{\beta}_0, \mathbf{H}) \tag{4}$$

and

$$\begin{aligned}
p(\boldsymbol{\alpha}) &\sim \mathcal{N}_N(\boldsymbol{\alpha}_0, \mathbf{H}_\alpha), \\
p(\boldsymbol{\eta}) &\sim \mathcal{N}_T(\boldsymbol{\eta}_0, \mathbf{H}_\eta).
\end{aligned} \tag{5}$$

In particular, if we assume state effects, then $\boldsymbol{\eta} = 0$ with probability one, which means assuming a degenerate prior for $\boldsymbol{\eta}$, with $\boldsymbol{\eta}_0 = 0$, $\mathbf{H}_\eta = \mathbf{0}$, the null matrix. Similarly, the model with time effects only is obtained by taking $\boldsymbol{\alpha}_0 = 0$, $\mathbf{H}_\alpha = \mathbf{0}$. The specification of the priors for $\boldsymbol{\Psi}$ and $\boldsymbol{\Phi}$ in the unstructured and structured cases will be detailed in the following sections. As is well known, within the Bayesian framework, the fixed and random effects differ only in that the random effects require one more level in the hierarchy. Therefore, in the case of random effects, we must also specify a prior distribution for the (vector valued) hyperparameters $\boldsymbol{\alpha}_0, \boldsymbol{\eta}_0$ while all other prior distributions remain the same:

$$\begin{aligned}
p(\boldsymbol{\alpha} \mid \boldsymbol{\alpha}_0, \mathbf{H}_\alpha) &\sim \mathcal{N}_N(\boldsymbol{\alpha}_0, \mathbf{H}_\alpha), & p(\boldsymbol{\alpha}_0) &\sim \mathcal{N}_N(\boldsymbol{\alpha}_{00}, \mathbf{H}_\alpha^0), \\
p(\boldsymbol{\eta} \mid \boldsymbol{\eta}_0, \mathbf{H}_\eta) &\sim \mathcal{N}_T(\boldsymbol{\eta}_0, \mathbf{H}_\eta), & p(\boldsymbol{\eta}_0) &\sim \mathcal{N}_T(\boldsymbol{\eta}_{00}, \mathbf{H}_\eta^0).
\end{aligned} \tag{6}$$

Finally, we set $\mathbf{H}_\eta^{-1} \mid \nu_\eta, \mathbf{A}_\eta \sim \mathcal{W}_T(\nu_\eta, \mathbf{A}_\eta)$ and $\mathbf{H}_\alpha^{-1} \mid \nu_\alpha, \mathbf{A}_\alpha \sim \mathcal{W}_N(\nu_\alpha, \mathbf{A}_\alpha)$ when the model includes time or space random effects, where $\nu_\eta > T+1$, $\nu_\alpha > N+1$, $\mathbf{A}_\eta, \mathbf{A}_\alpha$ are two symmetric p.d. matrices, and where $\mathcal{W}_k(a, \mathbf{A})$ indicates a $k$-dimensional Wishart distribution with parameters $a$ and $\mathbf{A}$.

## 2.1 Unstructured covariances

The only constraints imposed on the two matrices correspond to the definition of a covariance matrix (symmetry and p.d.). The priors $p(\boldsymbol{\Psi})$ and $p(\boldsymbol{\Phi})$ are matrix-valued probability distributions. A convenient choice is to let the distributions of $\boldsymbol{\Psi}$ and $\boldsymbol{\Phi}$ belong to conjugate families. In particular, we choose the inverse Wishart priors:

$$\boldsymbol{\Psi} \sim \mathcal{IW}_T(\gamma, \mathbf{A}),$$
$$\boldsymbol{\Phi} \sim \mathcal{IW}_N(\nu, \mathbf{B}),$$

where $\mathbf{A}$ and $\mathbf{B}$ are symmetric p.d. matrices, while $\gamma$ and $\nu$ are two scalars. We adopt the following parametrization: $\mathbf{E}\boldsymbol{\Psi} = \mathbf{A}/\gamma$ and $\mathbf{E}\boldsymbol{\Psi}^{-1} = \mathbf{A}^{-1}/(\gamma - T - 1)$, and, correspondingly, $\mathbf{E}\boldsymbol{\Phi} = \mathbf{B}/\nu$ and $\mathbf{E}\boldsymbol{\Phi}^{-1} = \mathbf{B}^{-1}/(\nu - N - 1)$; thus, $\gamma > T + 1$ and $\nu > N + 1$. Hence, the lower the shape parameters ($\nu - N - 1$ and $\gamma - T - 1$), the less informative the distribution.

One of the main drawbacks of the inverse Wishart distribution is that the uncertainty of the matrix is determined by the single scalar parameter ($\nu$ or $\gamma$). This limitation of the Wishart prior induced some authors to consider generalizations such as, for instance, the generalized inverse Wishart distribution (see, for instance, Brown et al. (1994)). This kind of generalization is not explored here, and will be the object of further research. Despite its limits, we find that the Wishart model performs quite well in our simulations, even compared to its competitors based on the *structured* specification. Although in principle Wishart probability density function is defined also for $\gamma \leq T+1$, the resulting estimates in this case are very unstable. A less informative choice for $\gamma$ and $\nu$ guaranteeing stability of the estimates, that is, $\gamma = T + 2$, $\nu = N + 2$, proves to work reasonably well even when sizes $N$ and $T$ are (moderately) unbalanced. However, when the difference between the time and space dimensions increases, the prior precision parameters may differ substantially. In particular, the prior precision relative to the covariance matrix with the highest dimension tends to increase and, as a consequence, the weight of the prior expected value on the posterior expected value tends to increase.

Besides the advocated robustness against misspecification of the covariance function, an advantage of the unspecified structure of the covariance matrices is that, similarly to the multifactor approach, it can account for strong forms of dependence. This is in particular important in the spatial dimension where shocks common to all regions at a given time can determine a persistent spatial dependence. According to the definition given in Chudik et al. (2009), the process $\{\varepsilon_{it}\}$ is cross-sectionally strongly dependent if, as $N \to \infty$,

$$N^{-1}\zeta_1(\boldsymbol{\Phi}) \geq K > 0$$

where $\zeta_i(\boldsymbol{\Phi})$ is the $i$th largest eigenvalue of matrix $\boldsymbol{\Phi}$. From Corollary 2.2 in Dette (2002), if $\nu = N + \kappa$ and $\mathbf{B} = \mathbf{I}_N$, then

$$\lim_N N\zeta_N(\boldsymbol{\Phi}^{-1}) = \lim_N N\left(1 - \sqrt{N/\nu}\right)^2 \leq \lim_N N\left(1 - N/\nu\right) = \kappa$$

and consequently,

$$N^{-1}\zeta_1(\boldsymbol{\Phi}) \geq 1/\kappa > 0.$$

Thus, the prior distribution for $\boldsymbol{\Phi}$ allows for strong cross-sectional (spatial) dependence.

## 2.2 Posterior distributions

The derivation of the conditional distributions is a simple matter because they are assumed to belong to conjugate families. For both the fixed and random effects models, the likelihood will be written as

$$
\begin{aligned}
L(\mathbf{y} \mid \boldsymbol{x}, \boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\eta}, \boldsymbol{\Psi}, \boldsymbol{\Phi}) \\
= \frac{|\boldsymbol{\Sigma}|^{-1/2}}{(2\pi)^{NT/2}} \exp\left\{ -\tfrac{1}{2}(\mathbf{y} - \boldsymbol{x}\boldsymbol{\beta} - D_s\boldsymbol{\alpha} - D_t\boldsymbol{\eta})'\boldsymbol{\Sigma}^{-1}(\mathbf{y} - \boldsymbol{x}\boldsymbol{\beta} - D_s\boldsymbol{\alpha} - D_t\boldsymbol{\eta}) \right\}
\end{aligned}
\tag{7}
$$

where $D_t = \mathbf{I}_T \otimes \boldsymbol{\imath}_N$ is the matrix of the time dummies and $D_s = \boldsymbol{\imath}_T \otimes \mathbf{I}_N$ is the matrix of the space dummies, and $\boldsymbol{\Sigma}$ as in (1) is equal to $\boldsymbol{\Psi} \otimes \boldsymbol{\Phi}$. The properties of Kronecker product allow us to compute $\boldsymbol{\Sigma}^{-1}$ as $\boldsymbol{\Psi}^{-1} \otimes \boldsymbol{\Phi}^{-1}$. Given the sets of priors of the previous subsection, we are able to write explicitly the posterior distributions for the fixed and random effects models.

For the fixed effects model the full conditional immediately writes:

$$
p(\mathbf{y}, \boldsymbol{x}, \boldsymbol{\beta}, \boldsymbol{\mu}, \boldsymbol{\Psi}, \boldsymbol{\Phi}) = p(\mathbf{y} \mid \boldsymbol{x}, \boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\eta}, \boldsymbol{\Psi}, \boldsymbol{\Phi})p(\boldsymbol{\beta})p(\boldsymbol{\alpha})p(\boldsymbol{\eta})p(\boldsymbol{\Psi})p(\boldsymbol{\Phi})
\tag{8}
$$

where

$$
p(\boldsymbol{\beta}) \propto |\mathbf{H}|^{-\frac{1}{2}} \exp\left( -\frac{1}{2}(\boldsymbol{\beta} - \boldsymbol{\beta}_0)'\mathbf{H}^{-1}(\boldsymbol{\beta} - \boldsymbol{\beta}_0) \right)
$$

and $p(\boldsymbol{\alpha})$ and $p(\boldsymbol{\eta})$ follow (5).

The posterior distribution of $\boldsymbol{\beta}$ and of the effects are both Gaussian. In particular,

$$
\boldsymbol{\beta} \mid \mathbf{y}, \boldsymbol{x}, \boldsymbol{\alpha}, \boldsymbol{\eta}, \boldsymbol{\Psi}, \boldsymbol{\Phi} \sim \mathcal{N}_k\left( \tilde{\boldsymbol{\beta}}, \tilde{\mathbf{H}} \right)
\tag{9}
$$

with

$$
\begin{aligned}
\tilde{\boldsymbol{\beta}} &= \mathbf{E}(\boldsymbol{\beta} \mid \mathbf{y}, \boldsymbol{x}, \boldsymbol{\alpha}, \boldsymbol{\eta}, \boldsymbol{\Psi}, \boldsymbol{\Phi}) = \tilde{\mathbf{H}}\left( \boldsymbol{x}'\boldsymbol{\Sigma}^{-1}(\mathbf{y} - D_s\boldsymbol{\alpha} - D_t\boldsymbol{\eta}) + \mathbf{H}^{-1}\boldsymbol{\beta}_0 \right), \\
\tilde{\mathbf{H}} &= \mathrm{var}(\boldsymbol{\beta} \mid \mathbf{y}, \boldsymbol{x}, \boldsymbol{\alpha}, \boldsymbol{\eta}, \boldsymbol{\Psi}, \boldsymbol{\Phi}) = \left( \boldsymbol{x}'\boldsymbol{\Sigma}^{-1}\boldsymbol{x} + \mathbf{H}^{-1} \right)^{-1}.
\end{aligned}
$$

The posterior distributions for $\boldsymbol{\alpha}$ and $\boldsymbol{\eta}$ are respectively:

$$
\begin{aligned}
\boldsymbol{\alpha} \mid \mathbf{y}, \boldsymbol{x}, \boldsymbol{\beta}, \boldsymbol{\eta}, \boldsymbol{\Psi}, \boldsymbol{\Phi} &\sim \mathcal{N}_N\left( \tilde{\boldsymbol{\alpha}}, \tilde{\mathbf{H}}_\alpha \right), \\
\tilde{\boldsymbol{\alpha}} = \tilde{\mathbf{H}}_\alpha\left[ D_s'\boldsymbol{\Sigma}^{-1}(\mathbf{y} - \boldsymbol{x}\boldsymbol{\beta} - D_t\boldsymbol{\eta}) + \mathbf{H}_\alpha^{-1}\boldsymbol{\alpha}_0 \right] &= \tilde{\mathbf{H}}_\alpha\left( \boldsymbol{\Phi}^{-1}\mathbf{M}_s\boldsymbol{\Psi}^{-1}\boldsymbol{\imath}_T + \mathbf{H}_\alpha^{-1}\boldsymbol{\alpha}_0 \right), \\
\tilde{\mathbf{H}}_\alpha = \left[ D_s'\boldsymbol{\Sigma}^{-1}D_s + \mathbf{H}_\alpha^{-1} \right]^{-1} &= \left[ \boldsymbol{\imath}_T'\boldsymbol{\Psi}^{-1}\boldsymbol{\imath}_T\boldsymbol{\Phi}^{-1} + \mathbf{H}_\alpha^{-1} \right]^{-1},
\end{aligned}
\tag{10}
$$

and

$$
\begin{aligned}
\boldsymbol{\eta} \mid \mathbf{y}, \boldsymbol{x}, \boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\Psi}, \boldsymbol{\Phi} &\sim \mathcal{N}_T\left( \tilde{\boldsymbol{\eta}}, \tilde{\Sigma}_\eta \right), \\
\tilde{\boldsymbol{\eta}} &= \tilde{\Sigma}_\eta\left( \boldsymbol{\Psi}^{-1}\mathbf{M}_t'\boldsymbol{\Phi}^{-1}\boldsymbol{\imath}_N + \mathbf{H}_\eta^{-1}\boldsymbol{\eta}_0 \right), \\
\tilde{\Sigma}_\eta &= \left[ \boldsymbol{\imath}_N'\boldsymbol{\Phi}^{-1}\boldsymbol{\imath}_N\boldsymbol{\Psi}^{-1} + \mathbf{H}_\eta^{-1} \right]^{-1}
\end{aligned}
\tag{11}
$$

where $M_s$ is the $N \times T$ matrix obtained by reshaping the vector $\mathbf{y} - \boldsymbol{x}\boldsymbol{\beta} - D_t\boldsymbol{\eta}$ such that the $N$ observations of year $t$ correspond to column $t$, $M_t$ is the $T \times N$ matrix obtained by reshaping the vector $\mathbf{y} - \boldsymbol{x}\boldsymbol{\beta} - D_s\boldsymbol{\alpha}$ such that the $T$ observations of state $i$ correspond to column $i$.

The posterior distributions of $\boldsymbol{\Phi}$ and $\boldsymbol{\Psi}$ are inverse Wishart:

$$
\begin{aligned}
\boldsymbol{\Psi} \mid \mathbf{y}, \boldsymbol{x}, \boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\eta}, \boldsymbol{\Phi} &\sim \mathcal{IW}(\gamma + N, \tilde{\mathbf{A}}) \quad \text{with } \tilde{\mathbf{A}} = \mathbf{U}'\boldsymbol{\Phi}^{-1}\mathbf{U} + \mathbf{A}, \\
\boldsymbol{\Phi} \mid \mathbf{y}, \boldsymbol{x}, \boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\eta}, \boldsymbol{\Psi} &\sim \mathcal{IW}(\nu + T, \tilde{\mathbf{B}}) \quad \text{with } \tilde{\mathbf{B}} = \mathbf{U}\boldsymbol{\Psi}^{-1}\mathbf{U}' + \mathbf{B},
\end{aligned}
\tag{12}
$$

where $\mathbf{U}$ is obtained by reshaping the vector of residuals $y - \boldsymbol{x}\boldsymbol{\beta} - D_s\boldsymbol{\alpha} - D_t\boldsymbol{\eta}$ into the $N \times T$ matrix, such that the $t$th column gives the $N$ residuals at time $t$.

In the case of a random effects model, the full conditional distribution includes one more component arising from the hierarchical step on the vectors $\boldsymbol{\eta}_0$ and $\boldsymbol{\alpha}_0$:

$$
\begin{aligned}
p(\mathbf{y}, \boldsymbol{x}, \boldsymbol{\beta}, \boldsymbol{\eta}, \boldsymbol{\alpha}, \boldsymbol{\eta}_0, \boldsymbol{\alpha}_0, \boldsymbol{\Psi}, \boldsymbol{\Phi}) &= L(\mathbf{y} \mid \boldsymbol{x}, \boldsymbol{\beta}, \boldsymbol{\eta}, \boldsymbol{\alpha}, \boldsymbol{\eta}_0, \boldsymbol{\alpha}_0, \boldsymbol{\Psi}, \boldsymbol{\Phi}) \\
&\times p(\boldsymbol{\beta})p(\boldsymbol{\eta} \mid \boldsymbol{\eta}_0)p(\boldsymbol{\eta}_0)p(\boldsymbol{\alpha} \mid \boldsymbol{\alpha}_0)p(\boldsymbol{\alpha}_0)p(\boldsymbol{\Psi})p(\boldsymbol{\Phi}).
\end{aligned}
\tag{13}
$$

Therefore, the posterior distributions of all the parameters remain the same, thus $\boldsymbol{\alpha}$ and $\boldsymbol{\eta}$ follow (10) and (11) and

$$
\begin{aligned}
\boldsymbol{\alpha}_0 \mid \mathbf{y}, \boldsymbol{x}, \boldsymbol{\beta}, \boldsymbol{\eta}, \boldsymbol{\eta}_0, \boldsymbol{\alpha}, \boldsymbol{\Psi}, \boldsymbol{\Phi} &\sim \mathcal{N}_N\left(\tilde{\boldsymbol{\alpha}}_0, \tilde{\mathbf{H}}_\alpha^0\right), \\
\tilde{\boldsymbol{\alpha}}_0 &= \tilde{\mathbf{H}}_\alpha^0\left[\mathbf{H}_\alpha^{-1}\boldsymbol{\alpha} + \mathbf{H}_\alpha^{0\,-1}\boldsymbol{\alpha}_{00}\right], \\
\tilde{\mathbf{H}}_\alpha &= \left[\mathbf{H}_\alpha^{-1} + \mathbf{H}_\alpha^{0\,-1}\right]^{-1},
\end{aligned}
$$

and similarly

$$
\begin{aligned}
\boldsymbol{\eta}_0 \mid \mathbf{y}, \boldsymbol{x}, \boldsymbol{\beta}, \boldsymbol{\eta}, \boldsymbol{\alpha}, \boldsymbol{\alpha}_0, \boldsymbol{\Psi}, \boldsymbol{\Phi} &\sim \mathcal{N}_N\left(\tilde{\boldsymbol{\eta}}_0, \tilde{\mathbf{H}}_\eta^0\right), \\
\tilde{\boldsymbol{\eta}}_0 &= \tilde{\mathbf{H}}_\eta^0\left[\mathbf{H}_\eta^{-1}\boldsymbol{\eta} + \mathbf{H}_\eta^{0\,-1}\boldsymbol{\eta}_{00}\right], \\
\tilde{\mathbf{H}}_\eta &= \left[\mathbf{H}_\eta^{-1} + \mathbf{H}_\eta^{0\,-1}\right]^{-1}.
\end{aligned}
$$

## 3 Structured covariances

Posterior estimation of unstructured covariances presented in Section 2.1 might reveal latent interactions and this can be advantageous when error correlation on the two dimensions, space and time, is not a straightforward function of spatial and temporal lags. On the other hand, there is a trade-off between robustness and parsimony of a model: under the separability assumption, the matrix $\boldsymbol{\Sigma}$ still depends on $(N(N + 1) + T(T + 1))/2$ parameters, in contrast with the models summarized in Table 1. When prior information on the spatial or the temporal interaction is available, modeling $\boldsymbol{\Psi}$ or $\boldsymbol{\Phi}$ as a known function of a fixed number of parameters could be preferable. In this section, we focus on the estimation of model (1) under this hypothesis. For the time dimension we consider an autoregressive specification, while for the matrix $\boldsymbol{\Phi}$ we define a matrix exponential specification. This is strictly related to

the matrix exponential spatial structure (MESS) models for the spatial errors proposed by LeSage and Pace (2007). For both parameters, we show that the posterior distribution is log-concave, permitting us to implement adaptive rejection sampling. Thus, unlike LeSage and Pace (2007), through a Bayesian approach, computations for MESS model estimation are feasible avoiding the finite series approximation of the exponential transformation.

## 3.1 Autoregressive model for $\Psi$

Starting from (3), again $\boldsymbol{\beta}$, $\boldsymbol{\eta}$ and $\boldsymbol{\alpha}$ follow (4) and (5), respectively. $\boldsymbol{\Psi}$ is now defined as a first order autoregressive matrix. Then, $|\boldsymbol{\Psi}| = (1 - \rho^2)^{T-1}$ and the inverse $\boldsymbol{\Psi}^{-1}$ can be written

$$\boldsymbol{\Psi}^{-1} = \frac{1}{1 - \rho^2} \left[ \mathbf{I}_T - \rho \mathbf{M}_2 + \rho^2 \mathbf{M}_3 \right] \tag{14}$$

where

$$\mathbf{M}_2 = \begin{pmatrix} 0 & 1 & \dots & 0 & 0 \\ 1 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 0 & 1 \\ 0 & 0 & \dots & 1 & 0 \end{pmatrix} \quad \text{and} \quad \mathbf{M}_3 = \begin{pmatrix} 0 & 0 & \dots & 0 & 0 \\ 0 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 1 & 0 \\ 0 & 0 & \dots & 0 & 0 \end{pmatrix}.$$

A reasonable assumption for $\rho$ can be a scaled Beta prior:

$$p(\rho) = \frac{\Gamma(c + d)}{\Gamma(c)\Gamma(d)} \left( \frac{\rho + a}{b - a} \right)^{c-1} \left( 1 - \frac{\rho + a}{b - a} \right)^{d-1} \tag{15}$$

with $-1 \leq a < b \leq 1$.

Except for the parameter $\rho$, the derivation of the posterior distributions of the other parameters is essentially the same as in Section 2.2.

The posterior of $\rho$ is a nonstandard distribution:

$$\begin{aligned} p(\rho \mid \mathbf{y}, \mathbf{X}, \boldsymbol{\beta}, \boldsymbol{\eta}, \boldsymbol{\alpha}, \Psi) \\ &\propto L(\rho \mid \mathbf{X}, \boldsymbol{\beta}, \boldsymbol{\eta}, \boldsymbol{\alpha}, \boldsymbol{\Psi}, \lambda) p(\rho) \\ &\propto (1 - \rho^2)^{N(T-1)/2} \left( \frac{\rho + a}{b - a} \right)^{c-1} \left( 1 - \frac{\rho + a}{b - a} \right)^{d-1} \exp^{-\frac{1}{2} \mathrm{tr} \Psi^{-1} C} \\ &\propto (1 - \rho^2)^{\frac{N(T-1)}{2}} \left( \frac{\rho + a}{b - a} \right)^{c-1} \left( 1 - \frac{\rho + a}{b - a} \right)^{d-1} \exp^{-\frac{1}{2} \frac{k_1 - k_2 \rho + k_3 \rho^2}{1 - \rho^2}}, \end{aligned} \tag{16}$$

where $\mathbf{C} = \mathbf{U} \boldsymbol{\Phi}^{-1} \mathbf{U}'$, $k_1 = \mathrm{tr}(\mathbf{C})$, $k_2 = \mathrm{tr}(\mathbf{M}_2 \mathbf{C})$ and $k_3 = \mathrm{tr}(\mathbf{M}_3 C)$. Log-concavity is easily verified (see Rowe (2002a)) and allows using adaptive rejection sampling as in Gilks and Wild (1992).

## 3.2 Matrix exponential $\boldsymbol{\Phi}$

The definition of a parametric spatial model usually stems from the knowledge of the spatial structure governing the process, often defined through an adjacency matrix $\mathbf{W}$. One of the leading examples is the spatial autoregressive covariance matrix $\boldsymbol{\Phi} = [(I - \lambda\mathbf{W})(I - \lambda\mathbf{W})']^{-1}$, also used in Zheng et al. (2008), where $\lambda$ is a scalar spatial correlation parameter and $\mathbf{W}$ is an adjacency matrix assumed to have non-negative elements and typically row-standardized such that $\|\mathbf{W}\|_\infty = 1$.

Here we propose a specification based on the matrix exponential: $\boldsymbol{\Phi} = \mathbf{e}^{\lambda\mathbf{W}}$, where for a square matrix $A$, $\mathbf{e}^A = \sum_{k=0}^\infty \frac{A^k}{k!}$. Chiu et al. (1996) proposed matrix exponential as a flexible way to specify an arbitrary covariance matrix. Some years before, Leonard and Hsu (1992) proposed to model the variance in a Bayesian hierarchical Gaussian model as the matrix exponential of an appropriately transformed Gaussian vector and showed advantages relative to the standard conjugate inverse Wishart prior distribution. The MESS model proposed by LeSage and Pace (2007) corresponds to $\boldsymbol{\Phi} = \mathbf{e}\{-\lambda\mathbf{W}\}\mathbf{e}\{-\lambda\mathbf{W}'\}$ and has the advantage of simplifying "the log-likelihood allowing a closed form solution to the problem of maximum-likelihood estimation". This occurs because of $\mathrm{tr}(\mathbf{W}) = 0$, that implies $\det(\boldsymbol{\Phi}) = \exp(\lambda\,\mathrm{tr}(\mathbf{W})) = 1$ for all $\lambda$. If $\mathbf{W}$ is a symmetric adjacency matrix, our specification corresponds to the covariance matrix of an MESS model, with a parameter $\lambda_{MESS} = -\lambda/2$.

Prior distributions of the parameters $(\boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\eta})$ are the same as in the above sections. As for $p(\boldsymbol{\Psi})$, it can be either the inverse Wishart as in Section 2.1 or the prior distribution $p(\rho)$ discussed in Section 3.1.

We assume for $\lambda$ a diffuse prior: $p(\lambda) \propto 1$. For both the fixed and random effects versions, the likelihood will be written as

$$
\begin{aligned}
&L(\mathbf{y} \mid \boldsymbol{x}, \boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\eta}, \boldsymbol{\Psi}, \lambda) \\
&\qquad = \frac{|\boldsymbol{\Sigma}|^{-1/2}}{(2\pi)^{NT/2}} \exp\left\{ -\tfrac{1}{2}(\mathbf{y} - \boldsymbol{x}\boldsymbol{\beta} - D_s\boldsymbol{\alpha} - D_t\boldsymbol{\eta})'\boldsymbol{\Sigma}^{-1}(\mathbf{y} - \boldsymbol{x}\boldsymbol{\beta} - D_s\boldsymbol{\alpha} - D_t\boldsymbol{\eta}) \right\}
\end{aligned}
\tag{17}
$$

where $\boldsymbol{\Sigma} = \boldsymbol{\Psi} \otimes \mathbf{e}^{\lambda W}$.

The posterior of $\lambda$ is a nonstandard distribution:

$$
\begin{aligned}
p(\lambda \mid \mathbf{y}, \mathbf{X}, \boldsymbol{\beta}, \boldsymbol{\eta}, \boldsymbol{\alpha}, \Psi) &\propto L(\mathbf{y} \mid \mathbf{X}, \boldsymbol{\beta}, \boldsymbol{\mu}, \boldsymbol{\Psi}, \lambda)p(\lambda) \\
&\propto \exp\left\{ -\mathrm{tr}\left(\frac{T\lambda W}{2}\right) - \frac{1}{2}\sum_{k=0}^\infty \frac{(-\lambda)^k}{k!}\mathrm{tr}\left(W^k\mathbf{U}\boldsymbol{\Psi}^{-1}\mathbf{U}'\right) \right\}
\end{aligned}
\tag{18}
$$

where the matrix $\mathbf{U}$ is defined in Section 2.2.

The log of the posterior distribution is easily seen to be concave, provided $\mathbf{W}$ is a symmetric matrix as shown in the Appendix, and adaptive rejection sampling as in Gilks and Wild (1992) can be implemented.

# 4  Computational aspects

We implemented a Gibbs sampling algorithm in Matlab, in order to obtain posterior estimates from the posterior distributions derived in the previous sections. Starting with initial values for the unknown parameters $\boldsymbol{\beta}$, $\boldsymbol{\alpha}$, $\boldsymbol{\eta}$, $\boldsymbol{\Psi}$ and $\boldsymbol{\Phi}$, say $\boldsymbol{\beta}_{(0)}$, $\boldsymbol{\alpha}_{(0)}$, $\boldsymbol{\eta}_{(0)}$, $\boldsymbol{\Psi}_{(0)}$ and $\boldsymbol{\Phi}_{(0)}$, the algorithm is based on the following steps (without lack of generality, we consider the fixed effects case):

**(a)** sample $\boldsymbol{\beta}_{(i)}$ from $p(\boldsymbol{\beta}_{(i)}|y, \boldsymbol{\alpha}_{(i-1)}, \boldsymbol{\eta}_{(i-1)}, \boldsymbol{\Psi}_{(i-1)}, \boldsymbol{\Phi}_{(i-1)})$, as in (9);

**(b)** sample $\boldsymbol{\alpha}_{(i)}$ from $p(\boldsymbol{\alpha}_{(i)}|y, \boldsymbol{\beta}_{(i-1)}, \boldsymbol{\eta}_{(i-1)}, \boldsymbol{\Psi}_{(i-1)}, \boldsymbol{\Phi}_{(i-1)})$, as in (10);

   sample $\boldsymbol{\eta}_{(i)}$ from $p(\boldsymbol{\eta}_{(i)}|y, \boldsymbol{\beta}_{(i-1)}, \boldsymbol{\alpha}_{(i-1)}, \boldsymbol{\Psi}_{(i-1)}, \boldsymbol{\Phi}_{(i-1)})$, as in (11);

**(c)** matrices $\boldsymbol{\Phi}$ and $\boldsymbol{\Psi}$ are sampled according to the four possible models:

   (i) *Unstructured model (Um):* no structure for both $\boldsymbol{\Psi}$ and $\boldsymbol{\Phi}$. The $[N(N+1)+T(T+1)]/2$ parameters are sampled from

$$p(\boldsymbol{\Phi}_{(i)}|y, \boldsymbol{\beta}_{(i-1)}, \boldsymbol{\alpha}_{(i-1)}, \boldsymbol{\eta}_{(i-1)}, \boldsymbol{\Psi}_{(i-1)})$$

   and

$$p(\boldsymbol{\Psi}_{(i)}|y, \boldsymbol{\beta}_{(i-1)}, \boldsymbol{\alpha}_{(i-1)}, \boldsymbol{\eta}_{(i-1)}, \boldsymbol{\Phi}_{(i-1)}),$$

   two inverse Wishart distributions with updated parameters as in (12);

   (ii) *Autoregressive model (AR):* autoregressive structure for $\boldsymbol{\Psi}$ and no structure for $\boldsymbol{\Phi}$. The $(1 + N(N+1)/2)$ parameters are sampled from (12) for

$$p(\boldsymbol{\Phi}_{(i)}|y, \boldsymbol{\beta}_{(i-1)}, \boldsymbol{\alpha}_{(i-1)}, \boldsymbol{\eta}_{(i-1)}, \rho_{(i-1)})$$

   and from (16) for $p(\rho_{(i)}|y, \boldsymbol{\beta}_{(i-1)}, \boldsymbol{\alpha}_{(i-1)}, \boldsymbol{\eta}_{(i-1)}, \boldsymbol{\Phi}_{(i-1)})$, through adaptive rejection sampling as in Gilks and Wild (1992) (ARS);

   (iii) *Expm model (Expm):* matrix exponential specification for $\boldsymbol{\Phi}$ and no structure for $\boldsymbol{\Psi}$. The $(1 + T(T+1)/2)$ parameters are sampled from (12) for

$$p(\boldsymbol{\Psi}_{(i)}|y, \boldsymbol{\beta}_{(i-1)}, \boldsymbol{\alpha}_{(i-1)}, \boldsymbol{\eta}_{(i-1)}, \lambda_{(i-1)})$$

   and from $p(\lambda_{(i)}|y, \boldsymbol{\beta}_{(i-1)}, \boldsymbol{\alpha}_{(i-1)}, \boldsymbol{\eta}_{(i-1)}, \boldsymbol{\Psi}_{(i-1)})$ as in (18), through ARS. Our specification for matrix exponential covariance differs from the MESS in that $\mathbf{W}$ is not necessarily an adjacency matrix[1]. In principle, the ARS algorithm does not need $\mathbf{W}$ to have zero entries in the main diagonal, but require a symmetric matrix, thus excluding row standardization in most of the cases.

---

[1]Recently, Rodrigues et al. (2014) highlighted some important failures of the matrix exponential spatial specification. If, in particular, $\mathbf{W}$ is a regular graph, the exponential matrix $\mathbf{e}^{\lambda \mathbf{W}}$ turns out to have a counterintuitive behavior: irrespective of the data, marginal and partial covariances between two regions have opposite sign. Similar difficulties in interpreting the spatial covariance structure, however, can arise using the SAR model (Wall (2004)), whenever the structure depends on an adjacency matrix.

(iv) *Fully structured model (AR-Expm):* autoregressive structure for $\boldsymbol{\Psi}$ and matrix exponential specification for $\boldsymbol{\Phi}$. The 2 parameters ($\rho$ and $\lambda$) are estimated sampling from (16) and (18) through ARS.

All specifications converge after a relatively low number of iterations. However, a higher computational cost is associated to all models that include at least one structured covariance matrix because sampling from an inverse Wishart distribution is faster than sampling from either (16) and (18). Both previous posterior distributions require, in fact, adaptive rejection sampling, and in some examples, the adaptive rejection sampling algorithm took up to few tens of seconds to sample an acceptable value.

# 5   Application

We illustrate the performance of the model proposed using simulated data and two real data sets. Simulation studies are considered in order to assess the performance of the model and the sensitivity to prior assumptions. The first data set consists of a panel of 46 states of the United States for a period of 30 years, from 1963 to 1992, and estimates the demand equation of cigarette consumption as a function of cigarette price and per capita disposable income. This data set is taken from Baltagi (2008) and has also been used for illustration purposes in other studies. A second application consists of house price data from Province of Rome from 2005 to 2011 and studies the effect of economic and demographic change on house price in 120 municipalities.

## 5.1   Simulated data

We simulate different scenarios, considering different values of $T$ and $N$ (equal to 16, 49 and 100). Different values for $\boldsymbol{\beta}$, $\boldsymbol{\alpha}$ and $\boldsymbol{\eta}$ have been considered, ranging in the interval $[-10, 10]$. Temporal structure is assumed to be a first order autoregressive with values for $\rho$ in $(-0.6, -0.3, 0, 0.3, 0.6)$. Concerning spatial structure, we simulate a fictitious squared area composed of $N$ squared regions ($\sqrt{N} = 4, 7, 10$ for each side), and $W$ is the corresponding symmetric adjacency matrix. The matrix $\boldsymbol{\Phi}$ is the exponential matrix defined in Section 3.2, $\boldsymbol{\Phi} = \mathbf{e}^{\lambda \mathbf{W}}$. For each adjacency matrix, different values for $\lambda$ are considered $(-0.6, -0.3, 0, 0.3, 0.6)$. Different true spatial covariance matrices, as, for example, spatial autoregressive, are also explored. For each scenario the absence of fixed effects, presence of time or spatial fixed effects and the presence of time and spatial effects are considered. For all the possible scenarios, we estimated the matrices $\boldsymbol{\Psi}$ and $\boldsymbol{\Phi}$ according to the four possible specifications (*Um, AR, Expm, AR-Expm*). For each model we run 1000 simulations, and the estimation is done by an MCMC algorithm, with 10000 MCMC iterations, after a burn in of 5000 iterations.

The estimates of $\boldsymbol{\beta}$, $\boldsymbol{\alpha}$ and $\boldsymbol{\eta}$ are stable under the different specification of the covariance matrix for all the values considered. For this reason, the table and figures that follow in this section refer to the case $\boldsymbol{\beta} = \boldsymbol{\alpha} = \boldsymbol{\eta} = 0$.

To evaluate the performance in estimating covariances, as a measure of distance between matrices, we use the normalized Frobenius norm of the difference between the true

$\lambda = -0.6$

|  | $\rho = -0.6$ | | | $\rho = 0$ | | | $\rho = 0.6$ | | |
|---|---|---|---|---|---|---|---|---|---|
| $\|\widehat{\boldsymbol{\Phi}} - \boldsymbol{\Phi}\|$ | **Um** | **AR** | **Expm** | **Um** | **AR** | **Expm** | **Um** | **AR** | **Expm** |
| $N = 16 \quad T = 16$ | 0.145 | 0.064 | 0.309 | 0.135 | 0.056 | 0.338 | 0.153 | 0.073 | 0.307 |
| $N = 16 \quad T = 100$ | 0.170 | 0.025 | 0.577 | 0.160 | 0.025 | 0.607 | 0.160 | 0.026 | 0.578 |
| $N = 100 \quad T = 16$ | 0.172 | 0.202 | 0.095 | 0.141 | 0.138 | 0.094 | 0.180 | 0.244 | 0.095 |
| $N = 100 \quad T = 100$ | 0.163 | 0.056 | 0.169 | 0.174 | 0.051 | 0.205 | 0.161 | 0.062 | 0.170 |
| $\|\widehat{\boldsymbol{\Psi}} - \boldsymbol{\Psi}\|$ | **Um** | **AR** | **Expm** | **Um** | **AR** | **Expm** | **Um** | **AR** | **Expm** |
| $N = 16 \quad T = 16$ | 0.181 | 0.508 | 0.056 | 0.074 | 0.003 | 0.084 | 0.185 | 0.509 | 0.069 |
| $N = 16 \quad T = 100$ | 0.164 | 0.477 | 0.164 | 0.225 | 0.224 | 0.241 | 0.158 | 0.347 | 0.160 |
| $N = 100 \quad T = 16$ | 0.226 | 0.955 | 0.021 | 0.042 | 0.547 | 0.030 | 0.218 | 0.376 | 0.022 |
| $N = 100 \quad T = 100$ | 0.171 | 0.426 | 0.052 | 0.095 | 0.117 | 0.075 | 0.169 | 0.541 | 0.052 |

$\lambda = 0.6$

|  | $\rho = -0.6$ | | | $\rho = 0$ | | | $\rho = 0.6$ | | |
|---|---|---|---|---|---|---|---|---|---|
| $\|\widehat{\boldsymbol{\Phi}} - \boldsymbol{\Phi}\|$ | **Um** | **AR** | **Expm** | **Um** | **AR** | **Expm** | **Um** | **AR** | **Expm** |
| $N = 16 \quad T = 16$ | 0.301 | 0.036 | 0.378 | 0.300 | 0.034 | 0.417 | 0.275 | 0.038 | 0.379 |
| $N = 16 \quad T = 100$ | 0.340 | 0.008 | 0.562 | 0.321 | 0.015 | 0.575 | 0.334 | 0.010 | 0.562 |
| $N = 100 \quad T = 16$ | 0.170 | 0.105 | 0.097 | 0.194 | 0.081 | 0.096 | 0.196 | 0.119 | 0.094 |
| $N = 100 \quad T = 100$ | 0.288 | 0.029 | 0.225 | 0.305 | 0.027 | 0.280 | 0.290 | 0.028 | 0.228 |
| $\|\widehat{\boldsymbol{\Psi}} - \boldsymbol{\Psi}\|$ | **Um** | **AR** | **Expm** | **Um** | **AR** | **Expm** | **Um** | **AR** | **Expm** |
| $N = 16 \quad T = 16$ | 0.169 | 0.490 | 0.105 | 0.145 | 0.004 | 0.170 | 0.131 | 0.492 | 0.100 |
| $N = 16 \quad T = 100$ | 0.244 | 0.477 | 0.382 | 0.344 | 0.232 | 0.574 | 0.246 | 0.350 | 0.391 |
| $N = 100 \quad T = 16$ | 0.195 | 0.971 | 0.022 | 0.078 | 0.474 | 0.031 | 0.201 | 0.371 | 0.021 |
| $N = 100 \quad T = 100$ | 0.165 | 0.430 | 0.067 | 0.171 | 0.117 | 0.121 | 0.166 | 0.559 | 0.070 |

Table 2: Frobenius norm of the difference $\widehat{\boldsymbol{\Phi}} - \boldsymbol{\Phi}$ and $\widehat{\boldsymbol{\Psi}} - \boldsymbol{\Psi}$ corresponding to different values of $\rho$ and $\lambda$ and different sample sizes.

and the estimated spatial and temporal matrices: $|\widehat{\boldsymbol{\Phi}} - \boldsymbol{\Phi}_0|/|\boldsymbol{\Phi}_0|$ and $|\widehat{\boldsymbol{\Psi}} - \boldsymbol{\Psi}_0|/|\boldsymbol{\Psi}_0|$, where $\boldsymbol{\Phi}_0$ and $\boldsymbol{\Psi}_0$ are respectively the true spatial and temporal covariances, $\widehat{\boldsymbol{\Phi}}$ and $\widehat{\boldsymbol{\Psi}}$ are respectively the estimated spatial and temporal covariances, and $|A| = \sqrt{\sum_{ij} |a_{ij}|^2}$.

Posterior estimates for $\rho$ are almost identical irrespective of whether the matrix $\boldsymbol{\Phi}$ has a Wishart prior or is a function of $\lambda$, and the same behavior is observed for $\lambda$ with respect to $\boldsymbol{\Psi}$. Thus, here we are only reporting results for the three models $Um$, $AR$ and $Expm$.

The unstructured model captures both temporal and spatial pattern as pointed out in Figures 1–4, where comparison of posterior distributions of matrices $\boldsymbol{\Phi}$ and $\boldsymbol{\Psi}$ with the true ones is shown for four scenarios.

Table 2 presents the Frobenius norm of the difference between the true and estimated covariance matrices (both $\boldsymbol{\Phi}$ and $\boldsymbol{\Psi}$), for small and large sample sizes and for some values of $\rho$ and $\lambda$. All matrices are standardized in order to have trace equal to their size. The matrix $\widehat{\boldsymbol{\Phi}}$ is the mean of the samples from the posterior Wishart distribution in the cases $Um$ and $AR$, while it is equal to the exponential matrix $\mathbf{e}^{\widehat{\lambda}\mathbf{W}}$ with $\widehat{\lambda}$ being the mean of the samples from the posterior distribution of $\lambda$ in the $Expm$ case. Similarly, the $\widehat{\boldsymbol{\Psi}}$ in
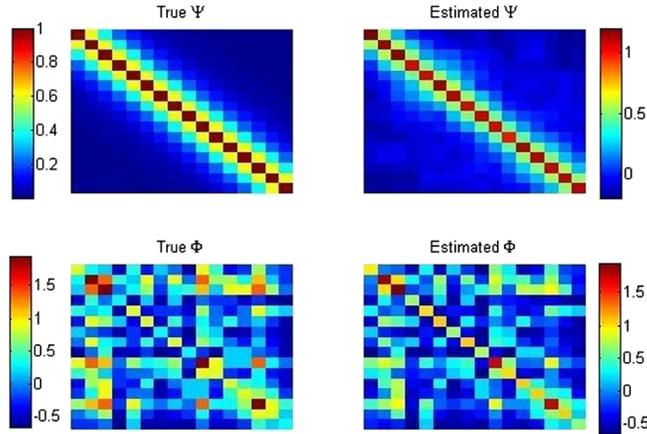
Figure 1: Posterior estimation of $\mathbf{\Phi}$ and $\mathbf{\Psi}$: unstructured model. $T = 16$, $N = 16$. The true $\mathbf{\Psi}$ matrix is AR(1) with parameter $\rho = 0.6$, the true $\mathbf{\Phi}$ is $Expm(\lambda \mathbf{W})$, with $\lambda = 0.6$.
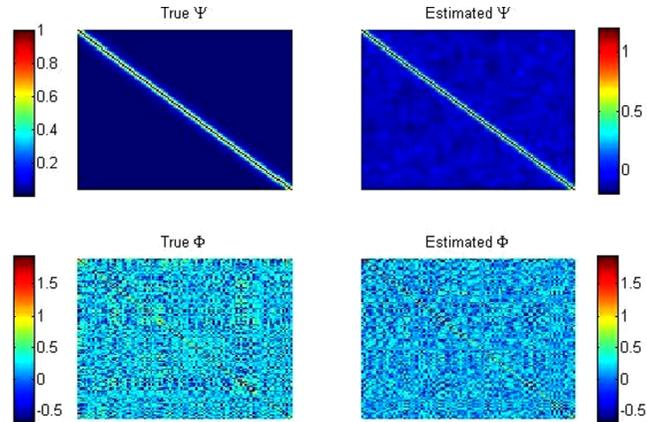


Figure 2: Posterior estimation of $\mathbf{\Phi}$ and $\mathbf{\Psi}$: unstructured model. $T = 100$, $N = 100$. The true $\mathbf{\Psi}$ matrix is AR(1) with parameter $\rho = 0.6$, the true $\mathbf{\Phi}$ is $Expm(\lambda \mathbf{W})$, with $\lambda = -0.6$.

the column denoted as $AR$ is the autoregressive covariance matrix computed using the posterior estimates of $\rho$. Unstructured models perform generally well. When the model is estimated assuming either $AR$ or $Expm$, we observe that the estimate is generally worse. In particular, the estimate of the corresponding structured matrix (i.e., $\mathbf{\Psi}$ when $AR$ model is assumed and $\mathbf{\Phi}$ when $Expm$ model is assumed, respectively) is worse; and, unexpectedly, the estimate of the matrix that is still estimated without assuming any structure is better. In order to evaluate the magnitude of the Frobenius norm presented in Table 2, we simulate 10000 pairs of matrices from inverse Wishart distributions with
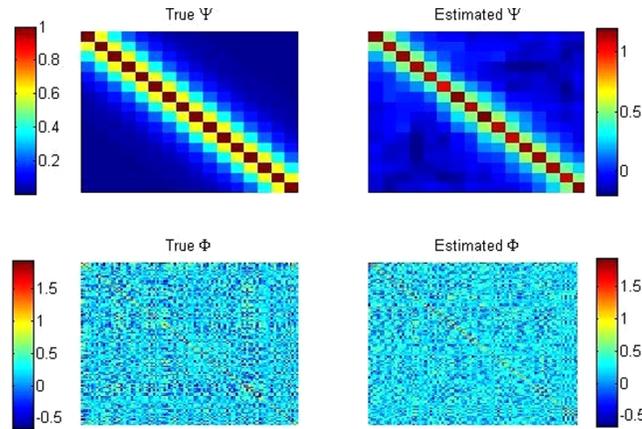
Figure 3: Posterior estimation of $\boldsymbol{\Phi}$ and $\boldsymbol{\Psi}$: unstructured model. $T = 16$, $N = 100$. The true $\Psi$ matrix is AR(1) with parameter $\rho = 0.6$, the true $\boldsymbol{\Phi}$ is $Expm(\lambda\mathbf{W})$, with $\lambda = -0.6$.
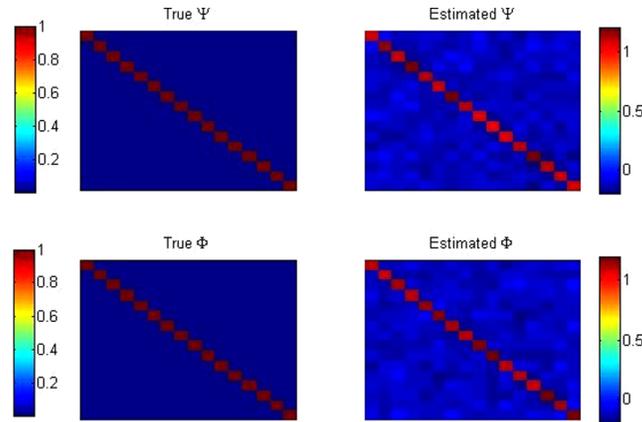


Figure 4: Posterior estimation of $\boldsymbol{\Phi}$ and $\boldsymbol{\Psi}$: unstructured model, $T = 16$, $N = 16$. The true $\boldsymbol{\Psi}$ matrix is $I_T$, the true $\boldsymbol{\Phi}$ is $I_N$.

expected values equal to the true matrices and shape parameters equal to 10. The values found in Table 2 result lower than the 2.5 percentile of the simulated results, except for the *Expm* and the *Um* estimates of $\boldsymbol{\Phi}$ when $N = 16$ and $\lambda = 0.6$; the latter, however, never exceeds the 10th percentile. In general, the structured estimates perform worse when the sample size ($N$ for $\boldsymbol{\Phi}$, $T$ for $\boldsymbol{\Psi}$) is small.

To confirm the fact that structured models are sensitive to the specification of the dependence pattern, we performed simulations from a data generating process with spatial covariance different from matrix exponential (we use an SAR with the same $\lambda$s).

Unsurprisingly, only unstructured models ($Um$ and $AR$ that are unstructured for $\boldsymbol{\Phi}$) are able to capture the spatial dependence correctly.

## 5.2  Cigarette demand data

The data set consists of a panel of 46 US states for a period of 30 years, from 1963 to 1992. The response variable is the real per capita sales of cigarettes by persons of smoking age (14 years and older) measured in packets per person on the log-scale (henceforth `cig`). Two explanatory variables are considered for constructing the demand equation: the average retail price of a packet of cigarettes measured in real terms on the log-scale (`p`) and the real per capita disposable income of each state on the log-scale (`inc`). Then, all the coefficients in the demand equation represent elasticities.

This data set has been widely used in spatial panel data literature (Baltagi and Levin (1986), Baltagi and Levin (1992), Elhorst (2005), Baltagi (2008), Lee and Yu (2010) among others). One of the challenges of this example is to take spatial dependence between neighboring states into account. Other challenges are in accounting for heterogeneities in space and time.

We run all the models described above, with and without temporal and spatial fixed/random effects. More details, as well as reasons to include state-specific and time-specific effects, are given in Baltagi (2008). The main findings of Baltagi and Levin (1986), Baltagi and Levin (1992) are a significant price elasticity and very small income elasticity (insignificant in Baltagi and Levin (1986) and significant but small in Baltagi and Levin (1992)). Also, the "bootlegging" effect is significant: the minimum price of neighboring states influences the cigarette consumption in a state.

Our estimated parameters $\boldsymbol{\beta}$ are in line with the results of Lee and Yu (2010): both the price and income elasticities are significant, opposite in sign. A negative significant price elasticity is also consistent with Baltagi and Levin (1986); income elasticity is smaller in absolute value, as in Lee and Yu (2010). The Baltagi and Levin (1986) analysis yields a significant negative effect of cigarette price on cigarette consumption, while there is no effect of income on cigarette consumption with an insignificant income elasticity.

Table 3 gives the results under different models and in Figure 5 posterior distribution of $\beta$ with and without other effects are shown. Table 3 shows that the effect of price is higher in absolute value when the exponential matrix model is assumed. Compared to the models presented in Elhorst (2010), which we do not report here, our estimates are more stable within the different models. To compare the different models, we compute the marginal likelihood drawn from the posterior distributions from Gibbs output according to Chib (1995) at each iteration and keeping the posterior mean values. These are reported in Table 3 together with the sum of squared errors (SSE). Both criteria suggest a better estimation arising from *unstructured* model.

Our model can also include spatially lagged regressors, which permit us to disentangle direct and indirect spatial effects. We do not report the results of the estimates obtained by adding lagged regressors in Table 3 since we want to stress the relative performance of unstructured and structured models. Moreover, by taking into account

| Model | p | 95% CI | inc | 95% CI | SSE | post likelihood |
|---|---|---|---|---|---|---|
| Um | | | | | | |
| | −0.4254 | [−0.4735, −0.3738] | 0.2863 | [0.2101, 0.3714] | 49.4580 | 8118.6 |
| $FE_{sp}$ | −0.4169 | [−0.4680, −0.3647] | 0.2738 | [0.1932, 0.3548] | 11.9912 | 8326.6 |
| $RE_{sp}$ | −0.4186 | [−0.4686, −0.3672] | 0.2686 | [0.1917, 0.3525] | 11.8692 | 8207.2 |
| $FE_{time}$ | −0.4353 | [−0.4902, −0.3786] | 0.3446 | [0.2495, 0.4355] | 38.4413 | 8413.7 |
| $RE_{time}$ | −0.4305 | [−0.4897, −0.3759] | 0.3558 | [0.2632, 0.4437] | 38.3616 | 8327.5 |
| $FE_{sp}\ FE_{time}$ | −0.4259 | [−0.4851, −0.3649] | 0.3324 | [0.2413, 0.4283] | 7.7115 | 8208.9 |
| $RE_{sp}\ RE_{time}$ | −0.4274 | [−0.4867, −0.3697] | 0.3321 | [0.2396, 0.4306] | 7.6937 | 8413.9 |
| AR | | | | | | |
| | −0.4552 | [−0.5407, −0.3665] | 0.1611 | [0.0526, 0.2779] | 52.4364 | 6264.7 |
| $FE_{sp}$ | −0.4381 | [−0.5237, −0.3532] | 0.1131 | [−0.0095, 0.2339] | 13.0839 | 6298.4 |
| $RE_{sp}$ | −0.4385 | [−0.5242, −0.3512] | 0.1153 | [−0.0009, 0.2325] | 13.1050 | 6285 |
| $FE_{time}$ | −0.4508 | [−0.5671,−0.3378] | 0.3355 | [0.1736, 0.4876] | 45.2281 | 6318.9 |
| $RE_{time}$ | −0.4543 | [−0.5705, −0.3368] | 0.3330 | [0.1706, 0.4942] | 45.2065 | 6297.7 |
| $FE_{sp}\ FE_{time}$ | −0.4272 | [−0.5486, −0.3083] | 0.2815 | [0.0954, 0.4731] | 8.9095 | 6285.6 |
| $RE_{sp}\ RE_{time}$ | −0.4283 | [−0.5505, −0.3073] | 0.2811 | [0.0913, 0.4711] | 8.8887 | 6319.2 |
| Expm | | | | | | |
| | −0.6497 | [−0.7517, −0.5465] | 0.2455 | [0.0900, 0.4307] | 50.9832 | 4984.7 |
| $FE_{sp}$ | −0.6270 | [−0.7321, −0.5247] | 0.2150 | [0.0407, 0.4360] | 12.2229 | 5139.4 |
| $RE_{sp}$ | −0.6280 | [−0.7357, −0.5260] | 0.2131 | [0.0490, 0.4236] | 12.1878 | 4999.8 |
| $FE_{time}$ | −0.6873 | [−0.8729, −0.5171] | 0.4741 | [0.3043, 0.6594] | 41.9080 | 5149.8 |
| $RE_{time}$ | −0.6945 | [−0.8708, −0.5242] | 0.4719 | [0.2964, 0.6528] | 41.8500 | 5136.7 |
| $FE_{sp}\ FE_{time}$ | −0.6410 | [−0.8165, −0.4680] | 0.5068 | [0.2897, 0.7160] | 7.9975 | 5002.4 |
| $RE_{sp}\ RE_{time}$ | −0.6448 | [−0.8092, −0.4809] | 0.4995 | [0.2889, 0.7231] | 8.0042 | 5145.8 |

Table 3: Posterior estimates of $\boldsymbol{\beta}$ under unstructured and structured models.

| Spatial lag models | SSPE | UM | SSPE |
|---|---|---|---|
| no FE | 9.4496 | no FE | 10.2582 |
| spatial FE | 4.3306 | spatial FE | 7.9736 |
| sp. lagged regress. | 12.7457 | sp. lagged regressors | 10.516 |
| sp. lagged regress. and spatial FE | 5.3404 | sp. lagged regress. and spatial FE | 8.4838 |

Table 4: SSPE of the prediction of the last 5 years of the panel for different spatial lag models and for different specifications of our unstructured model.

models with lagged regressors, we find that the performance in terms of prediction error is worsened, as shown in Table 4, which presents the sum of squared prediction errors (SSPE) of some of the spatial lag models of Table 1 and of our unstructured model.[2] In particular, the predictions of Table 4 are based on the subsample including the first 25 years. We also compared the out-of-sample prediction for some randomly chosen subsets of 6 states. The table presents four specifications of both the spatial lag model and the

---

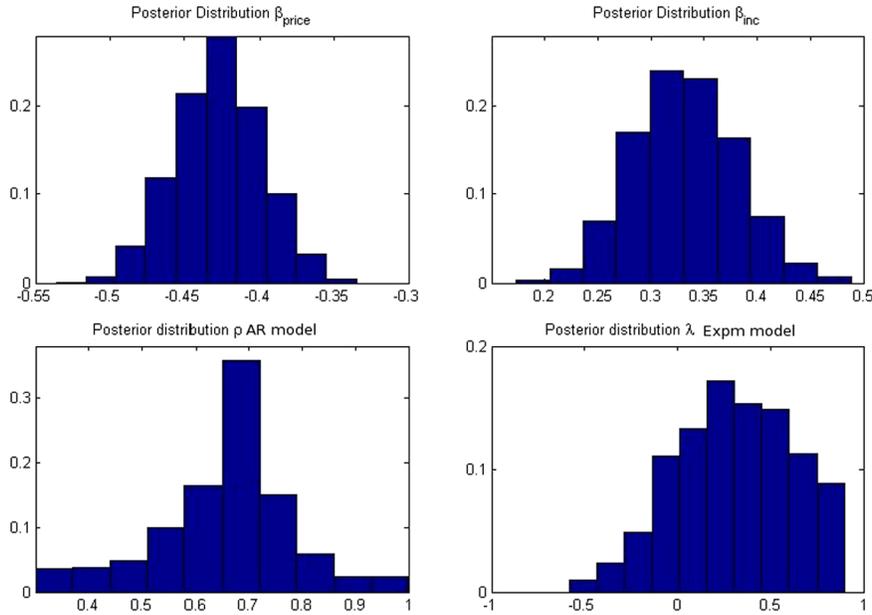[2]Here we limit ourselves to presenting the unstructured model that performs better, in general.

Figure 5: Cigarette Example: posterior distribution of $\beta$ under unstructured separable model (with spatial and temporal effects) and posterior distribution of $\rho$ and $\lambda$ under *AR* and *Expm* models (with spatial and temporal effects), respectively.

unstructured model, with or without spatial FE, and with or without spatially lagged regressors. Fixed effects mean spatial fixed effects in the case of prediction of future observations, and time fixed effects in the case of prediction of different locations. We find that adding spatially lagged regressors decreases the predictive accuracy of both our model and the spatial lag model, somehow confirming the findings of Elhorst (2005) and Lee and Yu (2010), who reported an insignificant bootlegging effect, unlike in Baltagi and Levin (1986). As noted by Lee and Yu (2010), these conflicting results might be explained by the inclusion of the spatial effects, which, if omitted, can lead to biased estimates. Their conclusion seems to be also validated by Table 4 showing that adding spatial fixed effects decreases the SSPE of all models considered.

The overall performance of our model is in line with that of spatial lag models. In the case of prediction in the time dimension, the spatial lag model seems to outperform the unstructured model: the best model is the spatial lag with fixed effects, while our best model is the one that includes, together with fixed effects, also the interaction between state and cigarette price (which is not reported in the table). In the case of prediction of out-of-sample states, the difference in SSPE clearly depends on the subset of states that are being predicted. However, in all the random subsets of locations we compared, the unstructured model appeared to perform at least slightly better than the spatial lag model. As in the case of prediction in the time dimension, adding spatially lagged regressors tends to increase the SSPE.
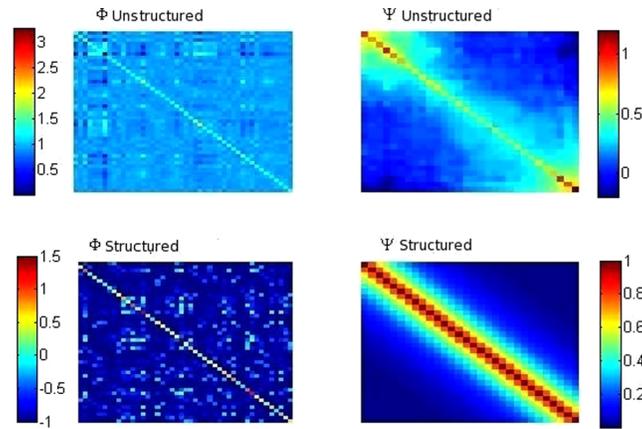
Figure 6: Cigarette Example: posterior estimation of temporal and spatial covariance matrices under structured and unstructured models, space and time fixed effects included.

Since our model is in the class of generalized spatial error models, the comparison of predictive accuracy (both in predicting out-of-sample years or locations) might help addressing the so called *Galton's problem*, that is, distinguishing spatial association in outcomes arising from common exposure to spatially correlated shocks from spatially dependent outcomes due to contagion (leading to the dichotomy *spatial lag vs spatial error*).

In Figure 6, posterior estimates of the two matrices $\boldsymbol{\Psi}$ and $\boldsymbol{\Phi}$ are shown, under unstructured (top matrices) and structured models (bottom matrices), including space and time fixed effects. Looking at the figure, the top matrices appear different from the bottom matrices, showing that structured and unstructured methods result in different patterns both in the space and in time dimension. Considering $AR$ model, running 5000 iterations (after a burn-in of 1000 iterations) from posterior distribution of $\rho$, a posterior estimate equal to 0.66 (95% credible interval [0.35, 0.928]) is obtained. The previous estimates are quite insensitive to the presence or absence of fixed or random effects, both spatial and temporal. On the other hand, observing the graphical representation of $Um$ posterior estimate of matrix $\boldsymbol{\Psi}$ under unstructured model, the temporal pattern shown is far from being first order autoregressive, suggesting the existence of three different regimes reciprocally uncorrelated, approximately corresponding to (1962–1971), (1972–1984) and (1985–1992). Our findings could be related to results found by Baltagi and Levin (1986) who pointed out that the temporal pattern between 1963–1970 is different from the following period, partly explained by the Fairness Doctrine Act effective before 1970 and the Impact of Taxation introduced in 1983. The two plots on the left in Figure 6, again, underline the advantage of unstructured model as respect to structured models. Considering, again, as a measure of distance between two matrices the normalized Frobenius norm, $Um$ posterior estimates of $\boldsymbol{\Phi}$ is 33% farther from the identity matrix compared to $Expm$ posterior estimate. The estimates of parameter $\lambda$

clearly depend on the choice of matrix $\mathbf{W}$, thus, a possible explanation of the difference between the two posterior estimates of the spatial correlation matrices could be due to incorrect specification of the weight matrix. Visual exploration of the posterior $Um$ matrix reveals the presence of correlation between regions not necessarily adjacent.

Zheng et al. (2008), through a Bayesian approach, found a positive spatial correlation, which decreases dramatically when temporal dependence is also included. Similarly, running $Expm$ model forcing $\boldsymbol{\Psi}$ to be the identity matrix, our posterior estimate of $\lambda$ increases and equals to 0.43 with a 95% credible interval [0.03, 0.81] compared to the estimate equal to 0.29 with 95% credible interval $[-0.312, 0.856]$, obtained with the $Expm$ model (with spatial and time fixed effects included and $\boldsymbol{\Psi}$ estimated unstructurally). Thus, our results suggest the advantage of considering simultaneously time and space dependence: by ignoring one source of variability, estimates of the other could be severely biased.

## 5.3   House price data

In this section, we apply our method to house price data from Italy. We use data on the average minimum and maximum price per square meter in the 120 municipalities of the Province of Rome, released by the *Agenzia dellle Entrate* (formerly *Agenzia del Territorio*), since 2005.

In housing data, spatial dependence is typically related to the so-called ripple effect: a great shock in the prices for houses in Rome propagates through neighboring towns. Without fixing *a priori* the weight matrix, we claim that our approach can capture the propagation phenomenon, not necessarily due to distances among observations or distance from Rome. Characteristics such as the commuting time and the existence of services (schools, pre-schools, hospitals) can be relevant to determine differences in the elasticities and in the correlations between house prices.

We adopt a simple model where house price drivers are the "classical" explanatory variables or fundamentals (see, among the others, Capozza et al. (2002)). The dependent variable $\log(Y_{it})$ is the logarithm of house price at time $t$ for municipality $i$. The regressors included are the following:

- `pop` population

- `income` average income, obtained from *Agenzia delle Entrate*,

- `car` number of cars over population, obtained from *Automobile Club d'Italia*,

- `business` number of firms over population, obtained from *Camera di Commercio di Roma*,

- `immigrants` number of immigrants over population, obtained from *Istituto Nazionale di Statistica, (ISTAT)*,

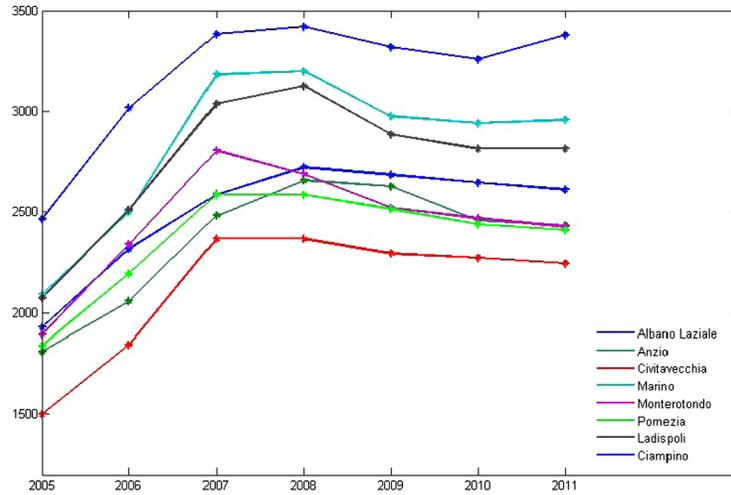- `elderly` percentage of population over 65 years old, obtained from *ISTAT*,
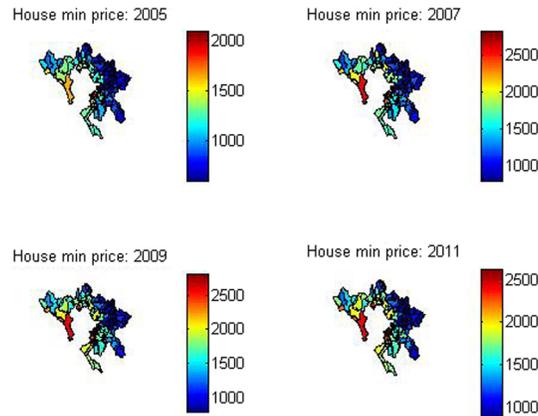
Figure 7: House price data for eight municipalities.



Figure 8: House price data for four time points.

- young percentage of population under 15 years old, obtained from *ISTAT*,

- school number of students enrolled in public school over population under 15, obtained from *Ministero dell'Istruzione*.

Data analysis is anticipated by a preliminary graphical investigation of spatial and temporal pattern. In Figure 7 and Figure 8, temporal and spatial pattern is shown, as an example, for a subsample of the data suggesting the necessity to consider correlations in both the dimensions. In many OECD countries, house prices in real terms rose sharply after the mid 1990s. As reported by Caliman and di Bella (2011), in the period 1999–2007
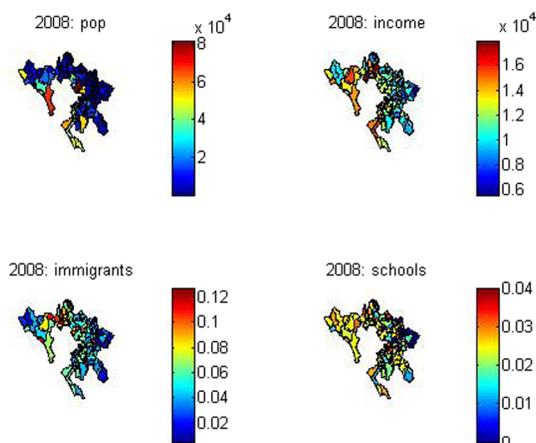
Figure 9: House price data: spatial distribution for four covariates.
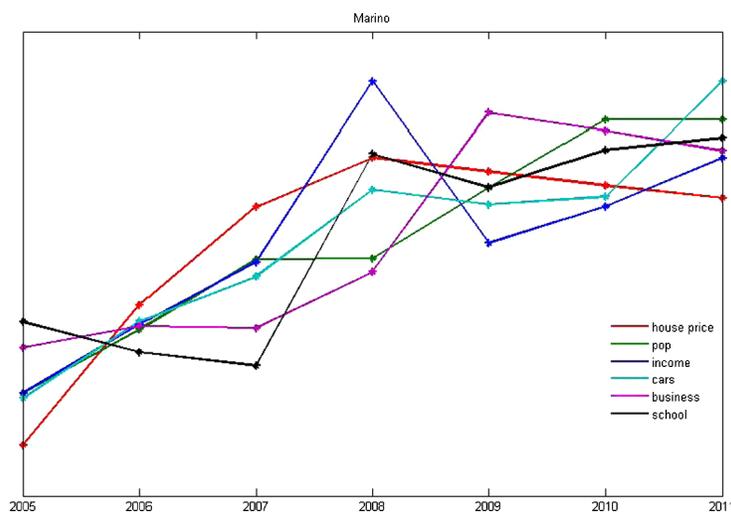


Figure 10: House price data: temporal distribution for five covariates in Marino.

house prices increased at a yearly average real rate of about 5% in the US, Euro Area and Canada. In Italy, in the same period (1999–2007) real house prices increased by 3.8% per year and the Affordability Ratio, that is, the ratio between house prices and incomes, showed a 20% increase. After 2008, in many areas in Italy, a slight decrease in house prices was observed followed by a plateaux. Figure 9 shows geographical distributions of four covariates (population, income, immigrants and school) for the year 2008. Figure 10 compares the time series of house prices and of some covariates (population, income, cars, business and school) for a selected municipality (Marino). Figure 9 and Figure 10,
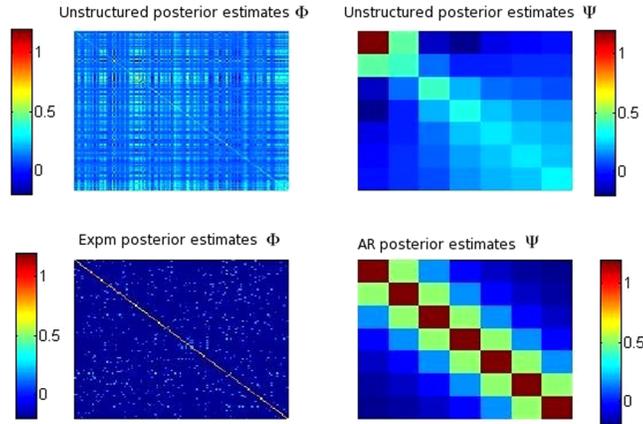
Figure 11: House price data: posterior estimation of matrices $\boldsymbol{\Psi}$ and $\boldsymbol{\Phi}$ under unstructured and structured models, space fixed effects included.

although referring to a subsample, suggest that, on the one hand, spatial association observed in house price distribution could be partially explained by spatial association of the covariates, while on the other hand, catching temporal pattern by means of covariates looks less straightforward.

We ran MCMC simulations for model *Um*, *AR*, *Expm* and *AR-Expm*, considering fixed or random, spatial, temporal and both spatial and temporal effects. Random and fixed effects produced essentially the same results. The best model in terms of comparison of observed and estimated values was the unstructured model with spatial fixed effects, while the computation of marginal likelihoods from Gibbs output, according to Chib (1995), suggested the choice of the model with time and space fixed effects.

Figure 11 compares posterior estimates of $\boldsymbol{\Phi}$ and $\boldsymbol{\Psi}$, with fixed spatial effects included, according to *Um* (top matrices) and *AR* and *Expm* models (bottom matrices). The two matrices in the first column look quite different, conveying the idea that the spatial pattern is unsatisfactorily captured by the adjacency matrix. The difference between the two matrices in the second column also suggests a possible misspecification of the first order autoregressive model for the temporal covariance. Posterior estimates of $\boldsymbol{\Psi}$ under *Um* model appear as a two-by-two block matrix. A latent shock may determine temporal pattern within the first three years and the last four years. Similar remarks arise from Figure 12, that compares posterior estimates of $\boldsymbol{\Phi}$ and $\boldsymbol{\Psi}$, without fixed effects.

The presence of spatial fixed effect modifies the posterior estimates of regression coefficients and covariance matrices, suggesting that the spatial pattern is explained by some latent factor not captured by the covariates. Figure 13 shows posterior distribution of the regression parameters under unstructured model with spatial fixed effects. Both $\beta_{pop}$ and $\beta_{car}$ are not significant when spatial fixed effects are included, while pop
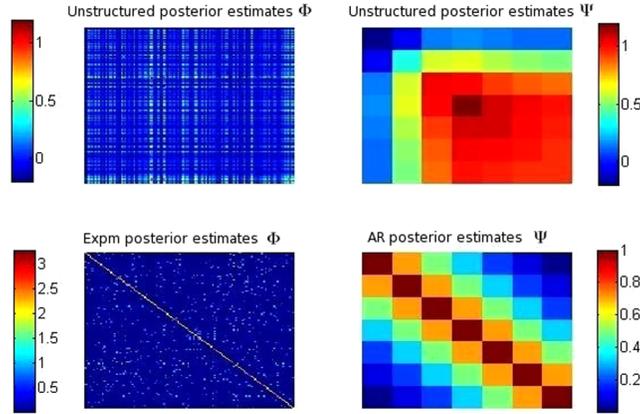
Figure 12: House price data: posterior estimation of matrices $\boldsymbol{\Psi}$ and $\boldsymbol{\Phi}$ under unstructured and structured models, without fixed effects.

results in a negative effect when no fixed effects are considered and a positive effect when only temporal fixed effects are taken into account; on the other hand, car has a significant coefficient only when no effects are considered. Furthermore, real incomes and percentage of students enrolled in local schools have a positive impact on house price that cannot be captured by spatial effects alone. Controversial is the positive effect found by the immigrants, contradicting the results in Caliman and di Bella (2011); generally migration decisions are affected by housing costs.

Repeating all the estimation by imposing the constraint $\boldsymbol{\Phi} = \mathbf{I}_N$ permits us to assess whether or not the data show a spatial correlation. A comparison between these new estimates with the previous ones (*Um* and *AR* only, since *Expm* model clearly can't be defined in this case) shows that estimates obtained without accounting for spatial correlation are quite different: the number of credible intervals not including the origin is larger, thus attesting possible bias due to misspecification of the constrained model. At the same time, credible intervals become larger which implies a contemporaneous loss in the accuracy of the estimation. For illustrative purposes, Table 5 reports the *Um* estimates of the model with no effects and time effects only.

The house price dynamic is negatively affected by interest rates (lower interest rates increase households' capacity to borrow). Since interest rates are constant among municipalities, we do not include this covariate, which would behave as a time-specific effect; however, eventual shocks on the interest rates in the period considered can contribute to explaining *Um* estimates of $\boldsymbol{\Psi}$.

We repeated the same comparison with spatial lag models performed in Section 5.2 to assess the relative predictive performances of our approach. If the focus is on predicting the house price for an arbitrary subset of the municipalities of the Province of Rome, our approach is dominated by the spatial lag model, in all specifications considered (with or without time effects, with or without spatially lagged regressors). This suggests that
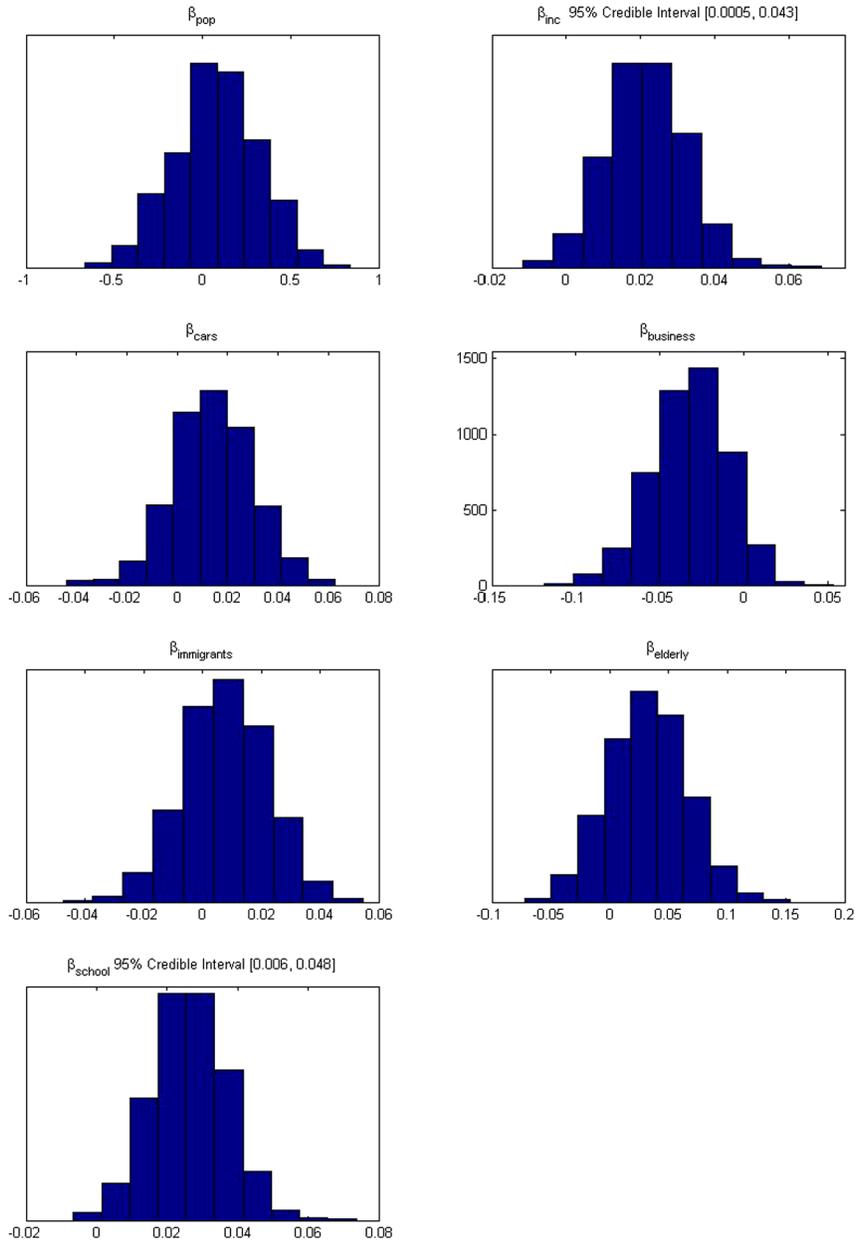
Figure 13: House price data: posterior estimation of $\boldsymbol{\beta}$.

a spatial error model, although flexible, is not appropriate to predict house price at different locations given characteristics of the municipalities because it is not able to adequately capture the contagion phenomena. The situation is more balanced when the

|  | $\mathbf{\Phi} = \mathbf{I}$ | | $\mathbf{\Phi} \sim \mathcal{IW}$ | |
|---|---|---|---|---|
| Variable | No Effects | 95% CI | No Effects | 95% CI |
| pop | 0.2457* | [0.1360, 0.3555] | 0.1199* | [0.0073, 0.2304] |
| inc | 0.0680* | [0.0323, 0.1053] | 0.0203* | [0.0006, 0.041] |
| cars | 0.0338 | [−0.0102, 0.0779] | 0.0238 | [−0.0033, 0.0532] |
| business | −0.0150 | [−0.0710, 0.0412] | −0.0297 | [−0.0651, 0.0044] |
| immigrants | 0.0304 | [−0.0104, 0.0729] | 0.0162 | [−0.0119, 0.0447] |
| elderly | −0.0774* | [−0.1496, −0.0048] | −0.0427 | [−0.0949, 0.0006] |
| school | 0.0235 | [−0.0117, 0.0589] | 0.0198 | [−0.0050, 0.0449] |
|  | $\mathbf{\Phi} = \mathbf{I}$ | | $\mathbf{\Phi} \sim \mathcal{IW}$ | |
| Variable | Time Effects | 95% CI | Time Effects | 95% CI |
| pop | 0.2536* | [0.1415, 0.3633] | 0.1496* | [0.0458, 0.2435] |
| inc | 0.0394* | [0.0069, 0.0728] | 0.0148 | [−0.0054, 0.0379] |
| cars | 0.0409 | [−0.0030, 0.0864] | 0.0272 | [−0.0041, 0.0580] |
| business | −0.0117 | [−0.0660, 0.0424] | −0.0271 | [−0.0631, 0.0109] |
| immigrants | 0.0367 | [−0.0088, 0.0824] | 0.0160 | [−0.0121; 0.0464] |
| elderly | −0.0721* | [−0.1440, −0.0025] | −0.0373 | [−0.0876; 0.0193] |
| school | 0.0213 | [−0.0123, 0.0554] | 0.0185 | [−0.0085; 0.0418] |

Table 5: *Um* Estimates and relative credible intervals of the models with no specific effects and with time effects, assuming a degenerate prior $\mathbf{\Phi} = \mathbf{I}$ (columns 1 and 2) and an inverse Wishart prior (columns 3 and 4). Parameters with * indicate a credible interval that does not contain zero.

final goal is to predict the house price in time: here in particular the unstructured model with spatial fixed effects reaches the lowest SSPE.

# 6  Sensitivity analysis

## 6.1  The choice of the hyperparameters

We conduct sensitivity analysis to investigate changes in posterior distribution by varying hyperparameter values.

The prior distribution for $\boldsymbol{\beta}$ is normal with mean $\boldsymbol{\beta}_0$ and variance matrix $\mathbf{H}$, with $\mathbf{H} = h_0 I$. A lower value for $h_0$ tends to shrink posterior estimates of $\boldsymbol{\beta}$ to $\boldsymbol{\beta}_0$, a reasonable value is $h_0 = 10^3$, thus reflecting poor prior information. The deviation of $\mathbf{H}$ from a scalar matrix does not affect posterior distribution of $\boldsymbol{\beta}$, provided that the trace remains the same. Similar prior variance matrices are chosen for the spatial and temporal effects. When random effect models are estimated, the prior variance for $\boldsymbol{\alpha}_0$ and $\boldsymbol{\eta}_0$ must be of the same order as that of $h_0$, otherwise effects are shrunk.

Increasing the value of $N$, the first component of (16) becomes too large and dominates the distribution, as an alternative to (15) the following prior distribution can be used, with $k$ approximately as large as $N$:

$$p(\rho) \propto (1 - \rho^2)^{-k}. \tag{19}$$

The choice of the shape hyperparameters in inverse Wishart prior distributions has always been critical, especially when it is desired to achieve arbitrarily high noninformativity (Gelman (2006)). When unstructured model is assumed, the posterior distributions of $\boldsymbol{\Psi}$ and $\boldsymbol{\Phi}$ seem, in fact, to be more sensitive to the order of magnitude of the shape parameters than to the form of the matrices $\mathbf{A}$ and $\mathbf{B}$. Identity matrices are, thus, assumed for $A$ and $B$. The smaller the parameters, the higher the uncertainty on the prior covariance matrix. However, the values of $\nu < N + 2$ and $\gamma < T + 2$ correspond to a prior distribution with no moments. In principle, the posterior distribution will have finite moments provided that the updated shape parameters $\nu + T$ and $\gamma + N$ satisfy $\nu + T \geq N + 2$ and $\gamma + N \geq T + 2$. However, choosing the values of $\nu$ and $\gamma$ below the critical thresholds $(N + 2, T + 2)$, will give unstable estimates of $\boldsymbol{\Phi}$ and $\boldsymbol{\Psi}$, that is why we set $\gamma = T + 2$ and $\nu = N + 2$.

From the cigarette data example we can observe that the posterior estimates of $\boldsymbol{\beta}$ are stable with respect to the choice of hyperparameters and the introduction of spatial or temporal effects, they change whenever a structured or unstructured model is assumed. Moreover, the posterior variance of $\boldsymbol{\beta}$ is larger when a structured model is assumed compared to an unstructured model.

To see how stable the final estimates are, multiple MCMC runs were conducted with different initial values and starting points.

## 6.2   The separability assumption

Statistical space–time modeling has been using separable covariance functions to a large extent because the structure of separable covariances reduces the number of parameters. Nonetheless, a well-known limitation of separable covariance functions is that they do not allow for space–time interactions in the covariance (see, e.g., Kyriakidis and Journel (1999), Cressie and Huang (1999)). For all these reasons, several papers have been devoted to test the reasonability of the hypothesis of separability of covariances; we cite Mitchell et al. (2006) among others. In many applications, a separable covariance structure can be seen as an approximation of a non-separable population covariance matrix. In all these cases, there is a trade-off between the gain related to dimension reduction and the loss in the accuracy of the estimation of the covariance matrix. In this light, Genton (2007), using real data on Irish wind speed, computed the nearest Kronecker approximation of a non-separable spatial–temporal covariance matrix in the Frobenius norm and found that the gain in the computation cost exceeded the loss in terms of mean squared prediction error.

In a similar spirit, in order to have a better understanding of how the lack of separability might affect our estimates, we focus on the sensitivity of our approach to the assumption of the separability of the spatial and temporal covariance components. We estimate the parameters of model (1) from a sample drawn from a spatial error model with a single regressor and a non-separable covariance matrix $\Sigma$. For the definition of the matrix, we follow the model introduced by Gneiting et al. (2007).

They proposed a general class of non-separable, stationary covariance function model for spatio-temporal random processes of the form
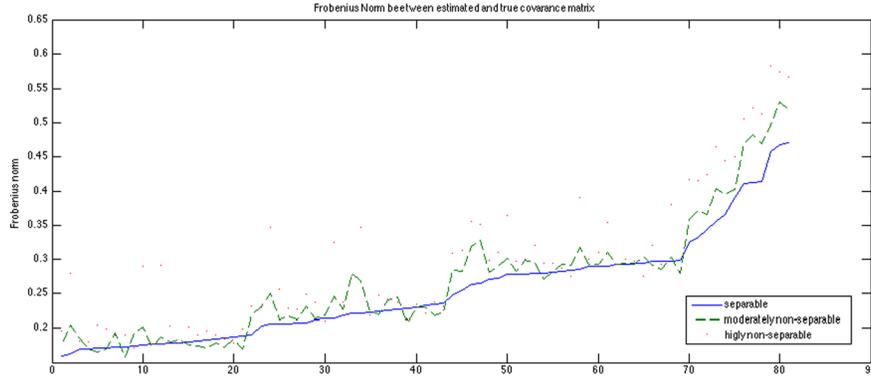
Figure 14: Frobenius norm of the distance between the true non separable matrix $\Sigma$ and the matrix $\widehat{\Psi} \otimes \widehat{\Phi}$ estimated assuming the unstructured conditionally separable model and for different choices of the parameters defining $\Sigma$.

$$C(h,u) = \begin{cases} \frac{1}{1+a|u|^{2\alpha}} & \text{if } h = 0, \\ \frac{1-\nu}{1+a|u|^{2\alpha}} \exp\left(\frac{-c||h||}{(1+a|u|^{2\alpha})^{\gamma/2}}\right) & \text{otherwise} \end{cases} \tag{20}$$

where $h$ is the distance between locations, $u$ is temporal lag, $a$ and $c$ are temporal and spatial nonnegative scale parameters, respectively, and $\alpha \in (0,1]$ is a smoothness parameter. The parameter $\gamma \in [0,1]$ controls the space–time interaction and $\gamma = 0$ yields a space–time separable correlation function for which the spatial correlations at different temporal lags $u$ are proportional to each other. We simulated different scenarios considering varying values for $a \in \{1, 2, 3\}$, $c \in \{0.1, 0.2, 0.3\}$, $\alpha \in \{0.1, 0.5, 1\}$, $\gamma \in \{0, 0.5, 1\}$ $\nu \in \{0.04, 0.5, 0.8\}$, and then estimated the matrix sigma, by assuming unstructured $\Psi$ and $\Phi$.

Figure 14 presents the Frobenius norm of the difference between the true and the estimated spatial and temporal matrices: $|\widehat{\Psi} \otimes \widehat{\Phi} - \Sigma|/|\Sigma|$, where $\widehat{\Phi}$ and $\widehat{\Psi}$ are respectively the estimated spatial and temporal covariances for each of the 243 models simulated: the three lines plotted (straight, dashed and dotted) represent the Frobenius norm for all the 81 different scenarios corresponding to a specific value of $\gamma$. In particular, the Frobenius norm does not seem to be much affected by the presence of a moderate non-separability ($\gamma = 0.5$), while there is a more relevant increase in the distance between the estimated and true covariance matrix in the case of maximum non-separability. However, the estimate of the $\beta$ parameter is not affected by any change in the parameter specifications, as Figure 15 shows.

# 7   Conclusion and discussion

We have proposed a Bayesian hierarchical model which incorporates both spatial and temporal correlation. Through flexibility of a Bayesian approach and thanks to the advances in MCMC computation, our proposed model has the potential to catch the
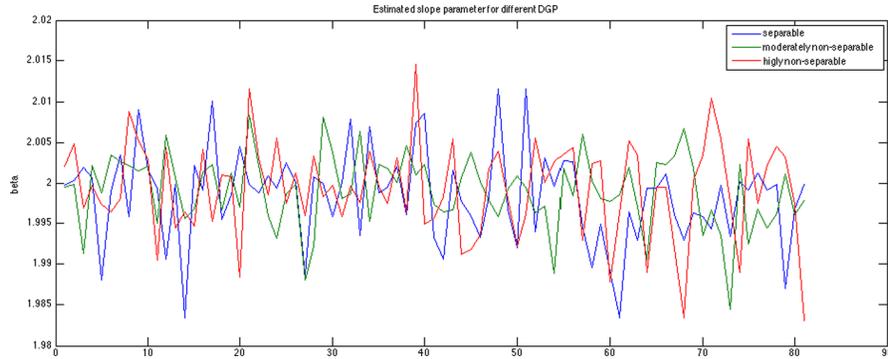
Figure 15: Posterior mean of $\hat{\beta}$ from the estimation of all 243 different spatial error models following (20) estimated assuming the unstructured conditionally separable model ($\beta = 2$).

eventual presence of hidden sources of correlation in both spatial and temporal dimensions. Previous contributions accounting for both temporal and spatial correlation have met some obstacles, which Bayesian approach can partially overcome, as already pointed out by Zheng et al. (2008).

The key feature of our model is the separability of the covariance matrix, which, besides the computational gain, allows for a direct interpretation of covariances as products of spatial and temporal components (see, for instance, Rowe (2002b)). The use of the Bayesian approach is convenient since it helps overcome the problem of the lack of identification of the matrices $\boldsymbol{\Phi}$ and $\boldsymbol{\Psi}$ by the introduction of the prior distributions.

By the dramatic reduction of the constraints imposed on covariance patterns, although increasing the number of parameters, we entail the whole class of SEMs (and of temporal error models) in a unique panel data model framework. Our *Um* model therefore lends itself to being the natural competitor for the recently developed multifactor approach: as for the multifactor error model, it is able to capture correlation structures whose nature is *a priori* unknown and allows for strong dependence. Unlike in the multifactor approach, the estimates of $\boldsymbol{\Phi}$ and $\boldsymbol{\Psi}$ obtained from the *Um* model open the way for a possible *ex-post* identification of the distance or adjacency matrix governing the spatial association. The model does not account for lagged variables (regressors or dependent variable), and so this feature needs to be developed further. However, spatially lagged regressors can be added easily to all specifications of our model as additional regressions in the equation.

From a computational viewpoint, the *AR* and *Expm* models offer a viable alternative to existing spatio-temporal models since they can both be estimated by implementing a simple ARS algorithm. However, the choice of a matrix exponential specification for the spatial covariance matrix has recently raised some concerns due to various rather problematic features of the matrix. Moreover, the link between the ma-

trix exponential and the more popular SAR models is not clear and deserves further exploration.

In all our models, the estimates of regressor coefficients (including the space or time specific effects) are rather robust; the *Expm* specification alone seems to affect the coefficient estimates in a significant way.

# Appendix A

## A.1 Log-concavity of posterior distribution of $\lambda$

$$\log p(\lambda \mid \mathbf{y}, \mathbf{X}, \boldsymbol{\beta}, \boldsymbol{\mu}, \Psi) \propto -\frac{1}{2} \sum_{k=0}^{\infty} \frac{(-\lambda)^k}{k!} \operatorname{tr}\left(W^k \mathbf{U} \boldsymbol{\Psi}^{-1} \mathbf{U}'\right) - \frac{1}{2} \operatorname{tr}(T\lambda \mathbf{W}).$$

Then,

$$
\begin{aligned}
\tfrac{\partial}{\partial \lambda} \log p(\lambda \mid \mathbf{y}, \mathbf{X}, \boldsymbol{\beta}, \boldsymbol{\mu}, \Psi) \quad &\propto -\tfrac{1}{2} \sum_{k=1}^{\infty} (-1)^k \operatorname{tr}\left(\mathbf{W}^k \mathbf{U} \boldsymbol{\Psi}^{-1} \mathbf{U}'\right) \tfrac{\lambda^{k-1}}{(k-1)!} - \tfrac{1}{2} \operatorname{tr}(T\mathbf{W}) \\
&= +\tfrac{1}{2} \operatorname{tr}\left(\mathbf{e}^{-\lambda \mathbf{W}} \mathbf{W} \mathbf{U} \boldsymbol{\Psi}^{-1} \mathbf{U}'\right) - \tfrac{1}{2} \operatorname{tr}(T\mathbf{W})
\end{aligned}
$$

while

$$
\begin{aligned}
\frac{\partial^2 \log p(\lambda \mid \mathbf{y}, \mathbf{X}, \boldsymbol{\beta}, \boldsymbol{\mu}, \Psi)}{(\partial \lambda)^2} &\propto -\frac{1}{2} \sum_{k=2}^{\infty} \operatorname{tr}\left(\mathbf{W}^k \mathbf{U} \boldsymbol{\Psi}^{-1} \mathbf{U}'\right) \frac{(-\lambda)^{k-2}}{(k-2)!} \\
&= -\frac{1}{2} \operatorname{tr}\left(\mathbf{W} \mathbf{e}^{\lambda \mathbf{W}} \mathbf{W} \mathbf{U} \boldsymbol{\Psi}^{-1} \mathbf{U}'\right) \\
&= -\frac{1}{2} \left((\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \boldsymbol{\eta} D_t - \boldsymbol{\alpha} D_s)'(\boldsymbol{\Psi}^{-1} \otimes \mathbf{W} \mathbf{e}^{-\lambda \mathbf{W}} \mathbf{W})(\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \boldsymbol{\eta} D_t - \boldsymbol{\alpha} D_s)\right)
\end{aligned}
$$

is nonpositive because both $\boldsymbol{\Psi}^{-1}$ and $\mathbf{W}\mathbf{e}^{-\lambda\mathbf{W}}\mathbf{W} = \mathbf{W}'\mathbf{e}^{-\lambda\mathbf{W}}\mathbf{W}$ are positive semidefinite.

# References

Anselin, L., Gallo, J. L., and Jayet, H. (2008). "Spatial panel econometrics." In: Mátyás, L. and Sevestre, P. (eds.), *The Econometrics of Panel Data*, volume 46 of *Advanced Studies in Theoretical and Applied Econometrics*, 625–660. Springer, Berlin, Heidelberg. 1037

Aston, J. A. and Gunn, R. N. (2005). "Statistical estimation with Kronecker products in positron emission tomography." *Linear Algebra and Its Applications*, 398: 25–36. Special Issue on Matrices and Mathematical Biology. MR2121342. doi: http://dx.doi.org/10.1016/j.laa.2003.11.018. 1036

Baltagi, B. H. (2008). *Econometric Analysis of Panel Data*. 1. John Wiley & Sons. 1047, 1051

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(1): 148–155.   1051, 1053, 1054

Baltagi, B. H. and Levin, D. (1992). "Cigarette taxation: Raising revenues and reducing consumption." *Structural Change and Economic Dynamics*, 3(2): 321–335.   1051

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(1): 5–51. Analysis of Spatially Dependent Data.   MR2395916. doi: http://dx.doi.org/10.1016/j.jeconom.2006.09.001.   1036, 1038

Baltagi, B. H., Song, S. H., and Koh, W. (2003). "Testing panel data regression models with spatial error correlation." *Journal of Econometrics*, 117(1): 123–150. MR2002299. doi: http://dx.doi.org/10.1016/S0304-4076(03)00120-9.   1037

Brown, P., Le, N., and Zideck, J. (1994). "Inference for a covariance matrix." In: Lindley, D., Smith, A., and Freeman, P. (eds.), *Aspects of Uncertainty: A Tribute to D. V. Lindley*, Wiley Series in Probability and Mathematical Statistics, 77–92. Wiley. MR1309689.   1041

Caliman, T. and di Bella, E. (2011). "Spatial autoregressive models for house price dynamics in Italy." *Economics Bulletin*, 31(2): 1837–1855.   1056, 1059

Capozza, D. R., Hendershott, P. H., Mack, C., and Mayer, C. J. (2002). "Determinants of real house price dynamics." NBER Working Papers 9262, National Bureau of Economic Research, Inc.   1055

Chib, S. (1995). "Marginal likelihood from the Gibbs output." *Journal of the American Statistical Association*, 90(432): 1313–1321. MR1379473.   1051, 1058

Chiu, T. Y., Leonard, T., and Tsui, K.-W. (1996). "The matrix-logarithmic covariance model." *Journal of the American Statistical Association*, 91(433): 198–210. MR1394074. doi: http://dx.doi.org/10.2307/2291396.   1045

Chudik, A., Pesaran, H., and Tosetti, E. (2009). "Weak and strong cross section dependence and estimation of large panels." Working Paper Series, European Central Bank 1100, European Central Bank.   1041

Cressie, N. and Huang, H.-C. (1999). "Classes of nonseparable, spatio-temporal stationary covariance functions." *Journal of the American Statistical Association*, 94(448): 1330–1340. MR1731494. doi: http://dx.doi.org/10.2307/2669946.   1036, 1062

Dette, H. (2002). "Strong approximation of eigenvalues of large dimensional Wishart matrices by roots of generalized Laguerre polynomials." *Journal of Approximation Theory*, 118(2): 290–304.   MR1932581. doi: http://dx.doi.org/10.1006/jath.2002.3725.   1041

Elhorst, P. J. (2005). "Unconditional maximum likelihood estimation of linear and log-linear dynamic models for spatial panels." *Geographical Analysis*, 37(1): 62–83.   1051, 1053

Elhorst, P. J. (2010). "Spatial panel data models." In: Fischer, M. M. and Getis, A. (eds.), *Handbook of Applied Spatial Analysis*, 377–407. Springer, Berlin, Heidelberg. 1035, 1037, 1051

Gelman, A. (2006). "Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper)." *Bayesian Analysis*, 1(3): 515–534. MR2221284. 1062

Genton, M. G. (2007). "Separable approximations of space-time covariance matrices." *Environmetrics*, 18(7): 681–695. MR2408938. doi: http://dx.doi.org/10.1002/env.854. 1036, 1062

Gilks, W. and Wild, P. (1992). "Adaptive rejection sampling for Gibbs sampling." *Applied Statistics*, 41(2): 337–348. 1044, 1045, 1046

Gneiting, T., Genton, M. G., and Guttorp, P. (2007). "Geostatistical space–time models, stationarity, separability and full symmetry." In: Finkenstädt, B., Isham, V., and Held, L. (eds.), *Monographs in Statistics and Applied Probability*. Chapman & Hall CRC Press. MR2307967. 1062

Kapoor, M., Kelejian, H. H., and Prucha, I. R. (2007). "Panel data models with spatially correlated error components." *Journal of Econometrics*, 140(1): 97–130. MR2395918. doi: http://dx.doi.org/10.1016/j.jeconom.2006.09.004. 1037

Kelejian, H. H. and Prucha, I. R. (2010). "Specification and estimation of spatial autoregressive models with autoregressive and heteroskedastic disturbances." *Journal of Econometrics*, 157: 53–67. MR2652278. doi: http://dx.doi.org/10.1016/j.jeconom.2009.10.025. 1037

Kyriakidis, P. and Journel, A. (1999). "Geostatistical space–time models: A review." *Mathematical Geology*, 31(6): 651–684. MR1694654. doi: http://dx.doi.org/10.1023/A:1007528426688. 1036, 1062

Lee, L.-F. and Yu, J. (2010). "Some recent developments in spatial panel data models." *Regional Science and Urban Economics*, 40(5): 255–271. 1051, 1053

Lee, L.-F. and Yu, J. (2012). "Spatial panels: Random components versus fixed effects." *International Economic Review*, 53(4): 1369–1412. MR2990371. doi: http://dx.doi.org/10.1111/j.1468-2354.2012.00724.x. 1037

Leonard, T. and Hsu, J. S. J. (1992). "Bayesian inference for a covariance matrix." *The Annals of Statistics*, 20(4): 1669–1696. MR1193308. doi: http://dx.doi.org/10.1214/aos/1176348885. 1045

LeSage, J. P. and Pace, R. K. (2007). "A matrix exponential spatial specification." *Journal of Econometrics*, 140(1): 190–214. MR2395921. doi: http://dx.doi.org/10.1016/j.jeconom.2006.09.007. 1044, 1045

Lindley, D. (1972). *Bayesian Statistics, A Review*. CBMS-NSF Regional Conference Series in Applied Mathematics. Society for Industrial and Applied Mathematics. MR0329081. 1039

Mitchell, M. W., Genton, M. G., and Gumpertz, M. L. (2006). "A likelihood ratio test for separability of covariances." *Journal of Multivariate Analysis*, 97(5): 1025–1043. MR2276147. doi: http://dx.doi.org/10.1016/j.jmva.2005.07.005. 1039, 1062

Naik, D. and Rao, S. (2001). "Analysis of multivariate repeated measures data with a Kronecker product structured covariance matrix." *Journal of Applied Statistics*, 28(1): 91–105. MR1834425. doi: http://dx.doi.org/10.1080/02664760120011626. 1036

Pesaran, M. H. (2004). "Estimation and Inference in Large Heterogeneous Panels with a Multifactor Error Structure." CESifo Working Paper Series 1331, CESifo Group Munich. 1040

Pesaran, M. H. and Tosetti, E. (2011). "Large panels with common factors and spatial correlation." *Journal of Econometrics*, 161(2): 182–202. MR2774936. doi: http://dx.doi.org/10.1016/j.jeconom.2010.12.003. 1040

Rodrigues, E., Assunçao, R., and Dey, D. K. (2014). "A closer look at the spatial exponential matrix specification." *Journal of Spatial Statistics*, 9: 109–121. MR3326834. doi: http://dx.doi.org/10.1016/j.spasta.2013.11.005. 1046

Rowe, D. B. (2002a). "Jointly distributed mean and mixing coefficients for Bayesian source separation using MCMC and ICM." *Monte Carlo Methods and Applications*, 8(4): 395–403. MR1943206. doi: http://dx.doi.org/10.1515/mcma.2002.8.4.395. 1044

Rowe, D. B. (2002b). *Multivariate Bayesian Statistics: Models for Source Separation and Signal Unmixing*. Taylor & Francis. 1064

Sarafidis, V. and Wansbeek, T. (2012). "Cross-sectional dependence in panel data analysis." *Econometric Reviews, Taylor & Francis Journals*, 31(5): 483–531. MR2903105. doi: http://dx.doi.org/10.1080/07474938.2011.611458. 1037

Shults, J. (2000). "Modeling the correlation structure of data that have multiple levels of association." *Communications in Statistics – Theory and Methods*, 29(5–6): 1005–1015. 1036

Srivastava, M., von Rosen, T., and von Rosen, D. (2008). "Models with a Kronecker product covariance structure: Estimation and testing." *Mathematical Methods of Statistics*, 17(4): 357–370. MR2483463. doi: http://dx.doi.org/10.3103/S1066530708040066. 1039

Wall, M. M. (2004). "A close look at the spatial structure implied by the CAR and SAR models." *Journal of Statistical Planning and Inference*, 121: 311–324. MR2038824. doi: http://dx.doi.org/10.1016/S0378-3758(03)00111-3. 1046

Zellner, A. (1962). "An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias." *Journal of the American Statistical Association*, 57(298): 348–368. MR0139235. 1039

Zheng, Y., Zhu, J., and Li, D. (2008). "Analyzing spatial panel data of cigarette demand: A Bayesian hierarchical modeling approach." *Journal of Data Science*, 6: 467–489. 1036, 1039, 1045, 1055, 1064

**Acknowledgments**