Regional Science and Urban Economics 64 (2017) 148–161

Contents lists available at ScienceDirect

Regional Science and Urban Economics

journal homepage: www.elsevier.com/locate/regsciurbeco

Embracing heterogeneity: the spatial autoregressive mixture model

Gary J. Cornwall*, Olivier Parent

University of Cincinnati, Cincinnati, OH 45221, United States

ARTICLE INFO

JEL classification:
C11
C21
R21

Keywords:
Mixture distributions
Spatial heterogeneity
Spatial models

ABSTRACT

In this paper a mixture distribution model is extended to include spatial dependence of the autoregressive type. The resulting model nests both spatial heterogeneity and spatial dependence as special cases. A data generation process is outlined that incorporates both a finite mixture of normal distributions and spatial dependence. Whether group assignment is completely random by nature or displays some locational "pattern", the proposed spatial-mix estimation procedure is always able to recover the true parameters. As an illustration, a basic hedonic price model is investigated that includes sub-groups of data with heterogeneous coefficients in addition to spatially clustered elements.

1. Introduction

Spatial models have long sought to control for two distinct aspects of the spatial paradigm, dependence and heterogeneity. Often the econometric tools are readily available to test and control for spatial dependence though heterogeneity has proved to be more challenging. Of the two, heterogeneity generally has a broader definition pertaining to varying structural parameters. Structural instability stemming from a spatial process is commonly designated as spatial heterogeneity and in order to obtain unbiased coefficient estimates it is necessary not only be aware of but also control for heterogeneity (Chamberlain, 1982). One proposal made to deal with this phenomenon is organizing the observations into well-delineated, often small clusters. If, in such a structure, the intercept is allowed to vary it can be noted that a form of spatial fixed effects has been implemented. It has been shown however that spatial fixed effects may in fact be spurious and not sufficient to correct for bias in the face of spatial dependence (Anselin and Arribas-Bel, 2013). In addition to fixed effects models, other solutions have been proposed including switching regression, random coefficient models, heteroskedasticity, amongst others. One alternative approach to modeling heterogeneity, often overlooked in the spatial context, is that of finite mixture models.

The continued increase in computational power over the last two decades has made mixture models a functional and efficient way to examine data with complex forms. Mixture models use discrete, latent variables as indicators which govern sub-groups of a population. A mixture of distributions is malleable enough to approximate most distributional forms and since the distributions in question are of known form this semi-parametric modeling technique is a viable alternative to other non-parametric methods. This approach emerged in economics via switching regression models (Quandt, 1972; Quandt and Ramsey, 1978) and has seen significant empirical use across many areas such as medical studies (Duan et al., 2007; Kottas et al., 2008), marketing (Allenby et al., 1998), finance (Kalli et al., 2013), management (Gabriel et al., 2014), and economics (Chotikapanich et al., 2007).

Mixtures of distributions appear when samples are drawn from a population comprised of several sub-groups, each of which is homogeneous within the sub-group population but heterogeneous between. The construction and inference in a Bayesian context is well-documented (see Escobar and West, 1995 and Frühwirth-Schnatter, 2006) though space is not typically seen as a confounder in such cases. These models are quite flexible and have the ability to adequately model data that otherwise would need to be examined non-parametrically (i.e. heavier-tailed and/or highly skewed data).

The general base of econometric literature pays little attention to the location of these subgroups and the potential impact on the relationships of interest. Gelfand et al. (2005) and Duan et al. (2007) introduce, in various forms, the Generalized Spatial Dirichlet Process model (GSDP) which accounts for dependency across space as a potential confounder with respect to medical research. A stick-breaking method is employed to determine the number of component distributions which are homogeneous via a base distribution (typically normal). This structure is employed again in Ji et al. (2009) though in all three cases the inference in an economic context is limited due to base level assumptions regarding the structure and impact of the spatial dependency.

Spatial dependency, as outlined by the aforementioned works, is a

* Corresponding author.
E-mail addresses: cornwagj@mail.uc.edu (G.J. Cornwall), parentor@ucmail.uc.edu (O. Parent).

http://dx.doi.org/10.1016/j.regsciurbeco.2017.03.004
Received 11 July 2016; Received in revised form 15 December 2016; Accepted 9 March 2017
Available online 10 March 2017
0166-0462/ © 2017 Elsevier B.V. All rights reserved.

latent process which is decomposed from the error term and exogenous to the overall data generation process (DGP). Furthermore, it is assumed that the distance between two observations is inversely related to the probability they are from the same distribution. Based on this assumption it is likely that individuals of a particular sub-group will tend to cluster across space. The implication being that group assignment is a deterministic function of location. Using this structure, the probability that the ith observation belongs to the gth sub-group is inferred from the posterior distribution rather than directly calculated (Duan et al., 2007).

Econometric applications of spatial dependency are noticeably different from the application mentioned above. The process is still exogenous to the model, however, in the case of a Spatial Autoregressive model (SAR), the dependency is observed via a filtered dependent variable. The focus has shifted in this context from decomposing the error term to understanding how the outcomes of individual agents spill over to their neighbors. The neighbor structure is imposed via a row-stochastic matrix rather than a direct measure of distance. This spatial weight matrix, W, allows for greater discretion on the part of a researcher in determining the relationships of interest and allows for a more fungible definition of space. More importantly, it allows for the calculation of partial derivatives to show how a change in the dependent variable is related to a change in explanatory variables coming from both one's own location and that of their neighbors. These effects, according to LeSage and Pace (2009), are labeled as direct and indirect effects.

This paper proposes a spatial mixture model which relaxes the assumptions made in the GSDP. The spatial autoregressive mixture model (hereafter referred to as 'SAR-M model') explicitly assigns a form to each of the component distributions and is not confined to the base distribution structure employed in the GSDP.[1] Following Aquaro et al. (2015) and LeSage and Chih (2016), a heterogeneous coefficient spatial autoregressive model is developed to allow for variation in the level of spatial dependence. However, the level of interaction between observations stays homogeneous within each group. Spillover effects from neighboring locations belonging to different groups will exert different impacts. To illustrate the flexibility of having heterogeneous spatial dependence in a cross sectional environment, a DGP is outlined which allows for neighbors to be of differing sub-groups within the mixture. Furthermore, a data augmentation step is employed to directly calculate the observation specific group probabilities. It will be shown that the SAR-M model provides unbiased estimates as opposed to its OLS counterpart and the SAR specification with no mixtures. Finally, unlike the GSDP the proposed model requires an explicit specification regarding the number of component distributions in the mixture.

The remainder of this paper will be organized as follows. Section 2 will outline the data generation process combining the mixture and spatial structures. Section 3 outlines the estimation method including prior structure, full joint posterior, and conditional parameter distributions. A sampling algorithm will be included in this section as well. Section 4 provides simulation parameters and Monte Carlo simulation results using the SAR-M model. Section 5 covers an empirical application in the form of hedonic pricing models. Section 6 summarizes the findings and offers avenues for additional research.

2. The spatial-mix model

The SAR specification relies on spillover effects between agents. Informally it can be stated that the outcomes of an agent rely not only on the attributes and decision of that particular agent, but also the outcomes of neighboring agents. There are a number of theoretical motivations for the observed correlation between nearby observations (see LeSage and Pace, 2009 for a full discussion). The DGP for this model is typically expressed in the form

$$y = \rho W y + x\beta + \epsilon, \quad \epsilon \sim N(0, \sigma^2 I_N) \tag{1}$$

where y denotes the N-dimensional vector of the observed dependent variable, x denotes the $N \times K$ matrix of exogenous explanatory variables, and β is the associated K-dimensional vector of parameters of interest. The $N \times 1$ vector of disturbance terms ϵ is assumed to be independently and identically distributed (i.i.d.) following a Normal distribution with variance $\sigma^2 I_N$. The $N \times N$ weight matrix, W, specifies the connectivity structure between observations. Each element w_{ij} will be different from zero if agents i and j are neighbors. This structure is exogenously defined by the researcher and can take a number of forms often resulting in a row-stochastic matrix. The value ρ is a scalar value that indicates the magnitude and direction of the spatial effect in the process. Since this weight matrix is row-normalized the value of ρ is bounded between $(1/\gamma_{min}, 1)$, where γ_{min} is the minimum eigenvalues of W.

It should be apparent that, in the absence of spatial dependence, the condition $\rho = 0$ will collapse (1) to the standard OLS process. If $\rho \neq 0$ then OLS is biased and inconsistent (see LeSage and Pace, 2009 for a full discussion). Recognize that this functional form is the most basic of spatial structures and little empirical work is conducted with such a structure.

Spatial dependence relies mainly on the assumption that clusters of individuals should, in principle, be based on geographical proximity. One aspect often overlooked is the considerable heterogeneity of behavior across individuals regardless of proximity. Spatial heterogeneity generally pertains to varying structural parameters over space. However, delineation over a well-defined group is often hard to assess. As detailed in Anselin and Arribas-Bel (2013), incorrect delineation might exacerbate spatially correlated and heteroskedastic error terms and create additional model misspecification. To account for unobserved heterogeneity, the observations are assumed to be drawn not from a single distribution but from a finite mixture of distributions. The introduction of spatial dependence in mixture models relaxes the assumption of an individual's distributional assignment being independent. In fact, the spatial proximity will generate correlation across individuals even if they happen to be of differing distributional assignment.

To relax the assumption of homogeneous spatial lag coefficients, Aquaro et al. (2015) and Lesage and Chih (2016) allow for heterogeneous coefficients for each spatial unit in a panel data setting.The proposed spatial mixture model assumes that spatial lag coefficients are only differentiated according to the group they belong to. Furthermore, while the above works are able to capture individual level effects (both spatial dependence and coefficients) it is only relevant in a panel data setting while the proposed model is cross-sectional in nature. Their work is, in the abstract, a corner case of that presented here such that N=G.

Data resulting from a mixture of distributions has a starkly different form than that of the spatial case and can be written in a number of ways. For the purposes of this discussion a general mixture DGP is used which allows for heterogeneous coefficients and variances across the sub-groups. As indicated previously a different functional form can be chosen for each of the components if necessary though that process will not be explicitly examined here. In practice, many distributional forms can be estimated strictly using a mixture of Gaussian distributions.

As described in Dempster et al. (1977), mixture models are often expressed as an incomplete data problem, where the missing data are represented by categorical latent variables indicating which mixture component generated each observation. To properly model mixtures of distributions, the following assumptions are needed:

Assumption 1. Each observation y_i, $(i = 1, \ldots, N)$ belongs to one of

[1] For expositional purposes all distributions in this paper are assumed to be normal though this is not a requirement.

g=1,...,G distinct groups. The number of groups, G, is known though an individual agent's group membership is not observed.

Assumption 2. The collection of observations which are assigned to each of the G groups are independent and identically distributed (i.i.d.) following a mixture of normal distributions.

Assumption 3. Each group is independent of the remaining groups such that the $G \times G$ variance-covariance matrix $\Sigma = diag(\sigma_1^2, \sigma_2^2, \ldots, \sigma_G^2)$.

Assumption 4. For each latent component indicator, $z_{ig} = 1$ if the ith observation is drawn from the gth component of the distribution and $z_{ig} = 0$ otherwise. Thus, each vector of component indicator $z_i = (z_{i1}, \ldots, z_{iG})$ is distributed according to a multinomial distribution consisting of one draw on G categories with probability $\pi = (\pi_1, \ldots, \pi_g)$.

Assumptions 1 to 3 allow for the observations $(y_1, \ldots, y_n)$ to form a sample from the following finite mixture of G Gaussian distributions:

$$p(y_i|x_i, \beta, \Sigma, \pi) = \sum_{g=1}^{G} \pi_g N(\tilde{y}_i|x_i\beta_g, \sigma_g^2), \qquad \sum_{g=1}^{G} \pi_g = 1 \tag{2}$$

where x_i is a $1 \times K$ vector of explanatory variables, $\beta_g = (\beta_{g1}, \ldots, \beta_{gK})'$ is a $K \times 1$ vector of coefficients specific to group g, and $\beta = (\beta'_1, \ldots, \beta'_G)'$ is a $KG \times 1$ vector of parameters of interest for all groups. The set of probabilities, $\pi = (\pi_1, \ldots, \pi_G)$ represents the mixture proportions such that $0 \le \pi_g \le 1$, and $\sum_{g=1}^{G} \pi_g = 1$. The non-spatial G-components univariate normal mixture distribution is obtained by setting $\tilde{y}_i = y_i$ and Assumption 3 guarantees that all observations are independent but have the same variance if they belong to the same group.

According to Assumption 1 the number of components, G, is to be fixed. For the purposes of this paper traditional model selection criteria (AIC, BIC, DIC, etc.) along with diagnostic plots are employed to determine the optimal number of components present in the mixture. If G is considered to be unknown and estimated as a parameter, a Dirichlet process can be implemented (see Gelfand et al., 2005 for such an example).

Being able to identify the component from which each observation, y_i, is generated requires a component-label indicator. To that end, Assumption 4 introduces a stochastic indicator, the role of which is to identify which group has generated each observation i=1,...,N. Let $z_i = (z_{i1}, \ldots, z_{iG})$ and $z = (z'_1, \ldots, z'_N)'$ be the $N \times G$ latent component indicator which identifies each observation i to a group g. Each vector of latent component z_i is conditionally independent distributed according to a Multinomial distribution with probability π.

The most common result of such a process is data which exhibits abnormalities in shape such as multiple modes and/or large tails. The number of identifiable modes should not be considered as strong evidence relevant to the size of G, that is a process with two modes could have two or more (less) component mixtures. One could consider abnormalities in the distribution to be weak evidence that a mixture model is an appropriate specification.

To extend the model outlined in (2) and incorporate the spatial dependence as described in (1), the parameters measuring the strength of spatial dependence are defined as $\psi = (\rho_1, \ldots, \rho_G)$ for each group g=1,...,G. In order to appropriately identify the observation with the appropriate group and spatial lag parameters, let $\widetilde{\psi} = z\psi'$, where $\widetilde{\psi}$ is an N-dimensional vector such that $\widetilde{\psi}_i = \rho_g$ if the observation i belongs to group g.

For the spatial autoregressive specification, it is assumed that $\tilde{y}_i = y_i - \widetilde{\psi}_i \sum_{j \in \delta_i} w_{ij} y_j$, where δ_i is defined to be the set of neighbors for agent i based on geographical proximity. Given the spatial weight matrix is row-normalized, $\widetilde{\psi}_i \sum_{j \in \delta_i} w_{ij} y_j$ represents the weighted average effect over the neighboring values of y_i. Note that, under this structure and set of assumptions, neighboring observations can identify with a different mixture component, g. The set of agents which belong to the mixture component g is defined as $I_g = \{i, z_{ig} = 1\}$ and it is assumed that each g is independent with distinct mean, $x_i\beta_g$ and variance, $\sigma^2{}_g$.

Assumption 5. In case of spatial dependence, Assumptions 2 and 3 hold only for $\tilde{y}_i$.

Given Assumptions 1 and 3 from the non-spatial case, it is easy to see that each observation y_i is no longer i.i.d. with the introduction of spatial dependence. Assumption 5 ensures that by setting $\tilde{y} = (I - \widetilde{\Psi} W)y$, $cov(\tilde{y}_i, \tilde{y}_j) = 0$ for any $i \neq j$.

Given the latent component indicator, the joint density of the data has the following representation

$$p(y|x, z, \psi, \beta, \Sigma, \pi) = |I_N - \widetilde{\Psi} W|^{-1} \prod_{g=1}^{G} \{(2\pi\sigma_g^2)^{N_g/2} \times \exp\left[-\frac{1}{2\sigma_g^2} \sum_{i \in I_g} \left(y_i - \widetilde{\psi}_i \sum_{j \in \delta_i} w_{ij} y_j - x_i \beta_g \right)^2 \right] \Bigg\}, \tag{3}$$

where $x = (x'_1, \ldots, x'_N)'$ is a N×K matrix of explanatory variables, $\beta = (\beta'_1, \ldots, \beta'_G)'$ is a $KG \times 1$ vector of parameters of interest, and $\widetilde{\Psi} = diag(\widetilde{\psi})$ is an N×N diagonal matrix for which each diagonal element $\widetilde{\psi}_i = \rho_g$ if the observation i belongs to group g and $N_g = \sum_{i \in I_g} z_{ig}$ represents the number of agents in each group g. As details in the next section, the latent component indicators z are independent multinomial and by including them with the observed dependent variable, the augmented model for (y,z) has the following full joint density

$$\begin{aligned} p(y, z|x, \beta, \Sigma, \psi, \pi) &= p(y|z, x, \beta, \Sigma, \psi, \pi) p(z|x, \beta, \Sigma, \psi, \pi) \\ &= \prod_{g=1}^{G} \prod_{i \in I_g} N(\tilde{y}_i|x_i\beta_g, \sigma_g^2) \pi_g^{N_g} \end{aligned} \tag{4}$$

Recognize that, as mentioned earlier, $\beta_g = (\beta_{g1}, \ldots, \beta_{gK})'$ is a $K \times 1$ vector of parameters specific to the gth group. Finally, note that marginalizing out the conditional likelihood defined in (3) over z will leave us with the spatial G- mixture model described at this outset of this section.

3. Estimation method

The likelihood function for models with mixture of normals poses highly complex computational challenges for any estimation procedure based on maximization (e.g. maximum likelihood). Within the parametric family and non-parametric settings, numerous approaches have been proposed to what constitutes a fascinating illustration of malleable approximations for this likelihood function (see Frühwirth-Schnatter, 2006 for a complete review). A flexible, parametric Bayesian framework is presented in this section following the seminal work of Escobar and West (1995).

In the case of analytically tractable distributions, as shown here, a modified Gibbs sampler can be employed to sample from the pertinent conditional distributions and standard inference as outlined by Escobar and West (1995) follows. An extension of the traditional Gibbs sampler proposed by LeSage and Pace (2009) is developed in this section. A Metropolis Hastings step is added for the spatial lag parameter by implementing a normal candidate distribution along with a tuned random-walk procedure. It is also possible, and perhaps slightly more efficient, to draw by inversion using a “griddy Gibbs” step as outlined by LeSage and Pace (2009).

Earlier it was assumed that the groups are i.i.d. and this is shown by establishing Σ to be the diagonal of the group variance-covariance matrix. This is convenient in that it allows for the model to collapse to a univariate draw mechanism within the sampler. The likelihood function represented in the posterior is given by (3) and priors are established for other parameters (ψ, Σ, β, π, z). It will be shown that, due to Assumption 4, it is possible to draw from a univariate distribution for $\tilde{y} = (I_N - \widetilde{\Psi} W)y$ rather than the more typical multivariate normal distribution for y. Drawing from the univariate case is merely done for convenience and computational efficiency.

A Normal-Inverted Gamma structure is used for the distributions of β and Σ respectively which implies the same structure will be seen in the posterior distribution. These priors are diffuse but proper as outlined by Geweke (2007) leading to posterior distributions which are also proper. A multinomial prior is established for the distribution of z and this can also be left uninformative with starting values equal to a uniform weight structure. A Dirichlet prior is used for the distribution of π, again a relatively uninformative approach can be taken. Different possibilities exist for the prior structure of $\psi = (\rho_1, \ldots, \rho_G)$ (see LeSage and Pace, 2009) though for the purposes of this paper the prior is assumed to follow a Beta distribution that takes the form of a relatively uniform distribution centered on a mean value of zero.

$$p(z_i|\pi) \sim M(1, \pi) \tag{5}$$

$$p(\pi) \sim D(\alpha_1, \ldots, \alpha_G) \tag{6}$$

$$p(\beta_g) \sim N(\beta_{0g}, V_{\beta_g}) \tag{7}$$

$$p(\sigma_g^2) \sim IG(a_g, b_g) \tag{8}$$

$$p(\rho_g) \sim Beta(d_0, d_0) \tag{9}$$

Using this prior structure and the likelihood function defined by (3) the posterior is of the form:

$$p(\beta, \Sigma, \psi|x, y, z) \propto p(y|x, \beta, \Sigma, \psi, z)p(z|\pi)p(\beta)p(\Sigma)p(\psi)p(\pi) \tag{10}$$

Given the latent indicator variable z, the data are classified for each iteration into G groups such that $x_g = \{x_i\}_{i\in I_g}$ and $\tilde{y}_g = \{\tilde{y}_i\}_{i\in I_g}$. Conditional on $\tilde{y}$, the posterior distributions will be independent across groups, as are their priors. As a result the sampler decouples into a set of G conditionally independent normal random samples. For each mixture component, g, the conditional distribution of β_g can be written as:

$$p(\beta_g|\sigma_g^2, \rho_g, x, y, z) \sim N(D_{\beta_g} d_{\beta_g}, D_{\beta_g}) \tag{11}$$

where

$$D_{\beta_g} = \left[\frac{1}{\sigma_g^2} x'_g x_g + V_{\beta_g}^{-1}\right]^{-1} \tag{12}$$

$$d_\beta = \frac{1}{\sigma_g^2} x'_g \tilde{y}_g + V_{\beta_g}^{-1}\beta_{0g} \tag{13}$$

$$\tilde{y} = (I_N - \tilde{\Psi} W)y \tag{14}$$

Turning to the estimation of the variance matrix $\Sigma = \text{diag}(\sigma_1^2, \sigma_2^2, \ldots, \sigma_G^2)$, each element σ_g^2 has a conjugate inverted Gamma prior $IG(a_g, b_g)$ and the posterior density of $p(\sigma_g^2|\beta_g, \rho_g, x, y, z)$ is generated from the following Inverted Gamma distribution,

$$p(\sigma_g^2|\beta_g, \rho_g, x, y, z) \sim IG(c_g, C_g) \tag{15}$$

where

$$C_g = a_g + \frac{N_g}{2} \tag{16}$$

$$c_g = b_g + \frac{1}{2}\epsilon'_g \epsilon_g \tag{17}$$

where $\epsilon_g = \tilde{y}_g - x_g\beta_g$.

The strength of spatial dependence $\psi = (\rho_1, \ldots, \rho_G)$ is estimated from the posterior distribution (18). Because it is not reducible to a standard distribution, the aforementioned Metropolis-Hastings step is incorporated into the MCMC sampling procedures. This step relies on a random walk proposal with normally distributed increments for each ρ_g, $g = 1\ldots, G$, such that $\rho_g^{new} = \rho_g^{old} + \eta_g N(0, 1)$ (see LeSage and Pace, 2009). The acceptance probability is calculated as the ratio of (18) evaluated at the old and new candidate draws. For each group g, the proposal tuning parameter η_g is systematically incremented or decremented when the acceptance rate moves below 0.40 or above 0.60, which results in an acceptance rate close to 0.50 after a burn-in period.

$$p(\rho_g|\rho_{-g}, \Sigma, \beta, x, y, z) = |I_N - \tilde{\Psi} W|(2\pi)^{-N_g/2} \prod_{g=1}^{G} \left\{(\sigma_g^2)^{-N_g/2} \exp\left[-\frac{1}{2\sigma_g^2}\epsilon'_g \epsilon_g\right]\right\} p(\rho_g), \tag{18}$$

where $p(\rho_g)$ is the Beta prior distribution.[2]

While z is a latent indicator, a step must be undertaken to calculate the probability that each observation belongs in every group. Since the dependent variable y is spatially correlated, the variance-covariance matrix is equivalent to $Var(y) = (I_N - \tilde{\Psi} W)^{-1}(z\Sigma z')(I_N - \tilde{\Psi} W)^{-1\prime}$. The simulation from a multivariate normal density is more complex and time consuming but not impractical. As previously explained, by leveraging the assumption of group independence it is possible to avoid drawing from the multivariate distribution and draw as though it were a non-spatial environment. As defined in (2), the spatial filter $\tilde{y}_i$ for each observation i is independent from any other observation. Therefore, for each observation, i, and a given group, g, the spatial filter is equivalent to $\tilde{y}_i = y_i - \rho_g \sum_{j\in\delta_i} w_{ij} y_j$, and the latent variable z_{ig} has the following conditional sample classification probability ω_{ig}:

$$\omega_{ig} \equiv Pr(z_{ig} = 1|y, x, \beta, \Sigma, \psi) \propto \pi_g N(\tilde{y}_i|x_i\beta_g, \sigma_g^2) \tag{19}$$

The $N\times G$ matrix ω is calculated by augmenting the model with predicted values of $\tilde{y}$ for each group g independently. For each group g and observation i, the predictive classification for each element $\tilde{y}$ is based on the following density

$$q_{ig} = \pi_g(2\pi\sigma_g^2)^{-1/2} \exp\left[-\frac{1}{2\sigma_g^2}\left(y_i - \rho_g \sum_{j\in\delta_i} w_{ij} y_j - x_i\beta_g\right)^2\right], \tag{20}$$

and is normalized in order to get the set of classification probabilities:

$$\omega_{ig} = \frac{q_{ig}}{\sum_{g=1}^{G} q_{ig}}. \tag{21}$$

The $N\times G$ matrix of new component indicator $z = (z'_1, \ldots, z'_N)'$ is obtained such that each G-dimensional row vector z_i is generated from an independent multinomial distribution:

$$p(z_i|y, x, \beta, \Sigma, \psi) \sim Mn(1, [\omega_{i1}\ \omega_{i2}\ \ldots\ \omega_{iG}]). \tag{22}$$

This holds only under the assumption that the groups themselves are independent of one another, as outlined in Assumption 4. Finally, the posterior distribution for the G-dimensional vector of probability π follows a Dirichlet distribution $p(\pi|\Sigma, \beta, \psi, x, y, z) \sim D(\alpha + N)$, where $\alpha = (\alpha_1, \ldots, \alpha_G)$ and $N = (N_1, \ldots, N_G)$.

It is apparent that if $\psi = 0$ then the spatial-mix DGP collapses to that of the mixture model outlined by (2). If z is an $(N \times 1)$ vector and $\psi = (\rho, \ldots, \rho)$ such that $\rho = \rho_1 = \rho_2 = \cdots = \rho_g$ then this structure collapses to the standard SAR form outlined in (1). Finally, if both of these conditions are true then the DGP collapses to the standard OLS framework. This is a flexible and general framework which nests the OLS, mixture, and SAR models as special cases and places no assumptions upon the geographical dispersion of the observations.

The full, modified Gibbs Sampler is as follows:

1. Classify x_g and $\tilde{y}_g$ into G groups
2. Draw from $p(\beta_g|\sigma_g^2, \rho_g, x, y, z) \sim N(D_{\beta_g} d_{\beta_g}, D_{\beta_g})$
3. Draw from $p(\sigma_g^2|\beta_g, \rho_g, x, y, z) \sim IG(c_g, C_g)$
4. Draw from $p(\rho_g|\rho_{-g}, \sigma_g^2, \beta_g, x, y, z) = |I_N - \tilde{\Psi} W|(2\pi)^{-N_g/2} \prod_{g=1}^{G} \left\{(\sigma_g^2)^{-N_g/2} \exp\left[-\frac{1}{2\sigma_g^2}\epsilon'_g \epsilon_g\right]\right\} p(\rho_g)$

[2] Note that ρ_g indicates the scalar level of spatial dependence for group g and that ρ_{-g} is designated as the scalar for each other group which is not g.

5. Draw from $p(z_i|y, x, \beta, \Sigma, \psi) \sim Mn(1, [\omega_{i1}\ \omega_{i2}\ \dots\ \omega_{iG}])$
6. Draw from $p(\pi|\Sigma, \beta, \psi, x, y, z) \sim D(\alpha + N)$
7. Return to 1 and iterate.

4. Monte Carlo simulations

To better understand the spatial mix DGP it has been rewritten below in reduced form. Sizing vectors ι_K and ι_G are defined to be vectors of ones with size $K \times 1$ and $G \times 1$ respectively. Again, W is considered to be an exogenous, row-stochastic matrix indicating the connectivity structure of each observation and $\widetilde{\Psi}$ is the $N\times N$ diagonal matrix described in (3). Using these definitions the DGP can be written in the reduced form

$$y = (I_N - \widetilde{\Psi}W)^{-1}\widetilde{X}\beta + (I_N - \widetilde{\Psi}W)^{-1}\tilde{\epsilon} \tag{23}$$

Where,

$$\widetilde{X} = (\iota'_G \otimes x) \odot (z \otimes \iota'_K) \tag{24}$$

$$\widetilde{\Psi} = \mathrm{diag}(z\psi') \tag{25}$$

$$\tilde{\epsilon} = (z\widetilde{\Sigma}^{1/2}) \odot \epsilon \tag{26}$$

$\epsilon \sim N(0, 1)$,

where $\widetilde{\Sigma}$ the $G \times 1$ vector of diagonal elements from Σ, $\otimes$ is the Kronecker product, and $\odot$ is the Hadamard product. The expanded $N\times KG$ matrix $\widetilde{X}$ of explanatory variables will be revisited after each interaction in the modified Gibbs sampler and because of this it is important to have a clear understanding of the mechanics involved. Consider the example of x which is a 5×2 defined below, with sample size N=5, the number of explanatory variables K=2 and the number of mixtures G=3. Conditional on the indicator matrix z, $\widetilde{X}$ is defined as:

$$x = \begin{matrix} 5 & 2 \\ 3 & 4 \\ 0 & 1 \\ 4 & 5 \\ 1 & 1 \end{matrix}, \quad z = \begin{matrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{matrix}, \quad \widetilde{X} = \begin{matrix} 0 & 0 & 5 & 2 & 0 & 0 \\ 3 & 4 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 4 & 5 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \end{matrix}$$

Note that the end result for $\widetilde{X}$ will be an $N \times KG$ group ordered matrix which, when multiplied by the vector β - of size $KG \times 1$ - produces the appropriate group effect on the outcome variable in question. A similar group ordered matrix for the variance is produced by (26). Furthermore, this structure allows for estimation of $\widetilde{\Psi}$ by allowing only the group specific scalar value on W to affect each agent within that group. To avoid label degeneracy, which is a trivial identification issue, the component weight vector is sorted in descending fashion upon completion of the estimation algorithm. Since the state labels typically lack any substantive interpretation and are merely used as a record keeping mechanism tracking the distribution, the posterior is permutation-invariant and priors must be proper (Geweke, 2007).

To test the efficacy of this modeling technique several different simulation environments were created in addition to a simple empirical application.

4.1. Signal to noise and simulation information

Following the advice and notation of Pace et al. (2011), special attention is paid to the signal-to-noise ratio (r^2) throughout the Monte Carlo simulations.[3] This section will outline simulation results in which the r^2 varies from high signal ($r^2 \approx 0.86$), medium-high signal ($r^2 \approx 0.62$), medium-low signal ($r^2 \approx 0.34$), and low signal ($r^2 \approx 0.17$). The calculation for this ratio can be found in (27) and the average value over the simulations will be reported in each table. Each simulation consists of 100 iterations through the DGP and modeling process with 5000 iterations through the sampler and a burn-in of 2000.

$$r^2 = 1 - \frac{\sigma_\epsilon^2\ \mathrm{tr}((I_N - \widetilde{\Psi}W)^{-2})}{\beta'\widetilde{X}'(I_N - \widetilde{\Psi}W)^{-2}\widetilde{X}\beta + \sigma_\epsilon^2\ \mathrm{tr}((I_N - \widetilde{\Psi}W)^{-2})} \tag{27}$$

The parameters for β are fixed over each simulation while the variance, σ_ϵ^2, is varied in order to obtain the desired level of signal. Sample size will be kept constant across the simulations and is reflective of a large sample where N=2000, meaning computational time for each model is approximately six minutes, the bulk of which is taken by calculating the determinate $|I_N - \widetilde{\Psi}W|$ during the M-H step of the sampler.[4] Finally, the draws for X are independent and Normally distributed with mean $\mu_X = 2$ and variance $\sigma_X^2 = 2$. It is important to note that, unlike traditional Monte Carlo simulations in which X is drawn from a distribution with mean zero, in order create some heterogeneity in the outcome the mean of X must be non-zero. The simulation data was created using Eq. (23) outlined earlier in this section.

The geographic location of each agent in the simulation is drawn from a normal distribution for both latitude and longitude. Simulations have been completed with varying degrees of spatial heterogeneity; from the completely heterogeneous case (Fig. 1a) to the independent locational structure (no heterogeneity). For brevity, the results presented here are restricted to a mixture of the two cases, half of the agents are clustered in a spatial heterogeneity form while the other half are permitted unrestricted locational options, see Fig. 1b. This structure is difficult in traditional modeling frameworks since delineation of the clusters is nearly impossible as there is no clear, visible pattern for which to draw the borders. Results are consistent throughout the domain of geographic assignment; from the fully segregated to random placement structures.

4.2. Homogenous spatial dependence

Before analyzing in detail the general spatial mixtures with heterogeneous spatial lag, particular attention is given to the importance of bias and group misspecification when spatial dependence is omitted from mixture models.[5] Group assignments, while randomized for each observation in the simulation, are kept constant in size across all simulations with $\pi = (0.45, 0.35, 0.20)$. In this particular case the DGP is forced into a homogeneous spatial level of spatial dependence where $\psi = (\rho_1, \rho_2, \rho_3) = (.7, .7, .7)$. For all simulations priors are kept relatively uninformative and set to: $\beta_{0g} = 0_K$, $V_{\beta_g} = 100I_K$, a_g=3, b_g=0.5, d_0=1.01 $\alpha_g = 2$, for all $g = (1\dots, G)$. Note that an OLS mixture model was also run on the simulated data in order to show that estimates are biased in the face of spatial dependence (see LeSage and Pace, 2009).

It is apparent from results presented in Table 1 that the SAR-M model provides estimates which are close to the true values and, as expected, the OLS estimates tend to be biased and inefficient. For each parameter the SAR-M model provides a distribution which is not only unbiased but exhibits less variance than that of the OLS mixture. It also can be seen that estimates of the variance present in the process are greatly inflated in the OLS mixture case relative to the SAR-M results. Finally, it is important to notice that the SAR-M model captures the appropriate level of variation in the process as indicated by the r^2 and R^2 values.

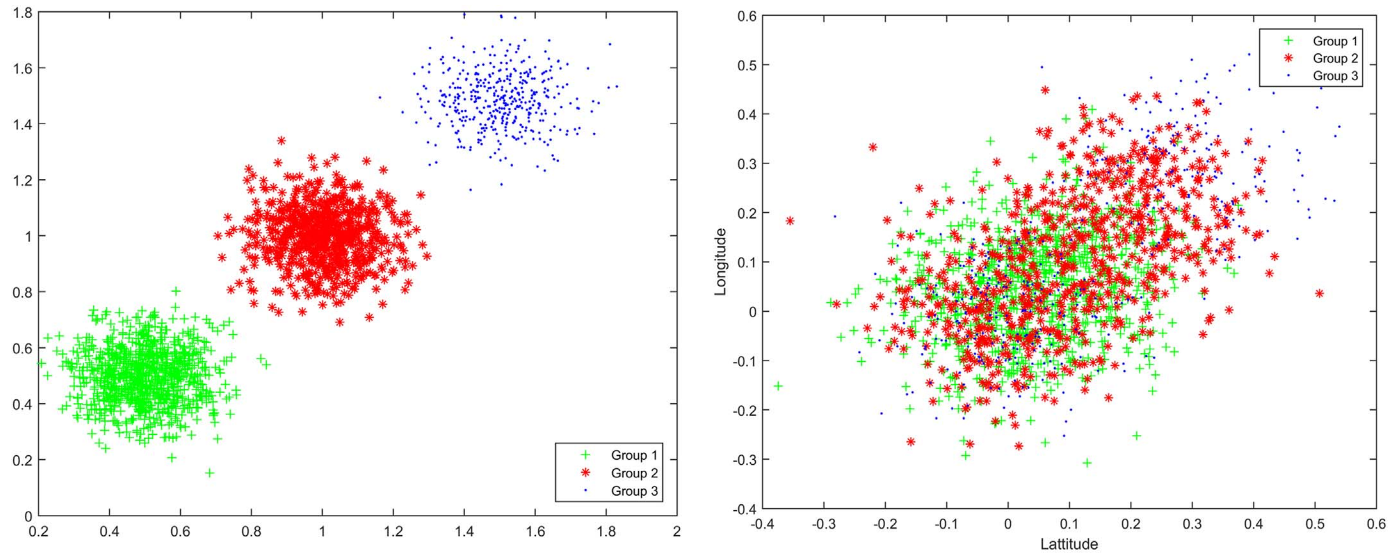
4.3. Simulation results

Results for the heterogenous spatial dependence mixture model are provided in Table 2 and show that the true parameters are recovered very well under high signal conditions. Not only are the posterior

[3] Note that while r^2 is used to denote the signal-to-noise ratio the notation R^2 refers to the traditional model based measure of captured variation.

[4] A special thanks to Dr. R. Kelley Pace and Dr. James LeSage for advice on increasing the computational efficiency for this process using a sparse incomplete LU factorization.

[5] Note that only one simulation is presented for the homogeneous spatial dependence case though the results are consistent across the domain of ρ.



(a) Complete Heterogeneity   (b) Quasi-independent Structure

Fig. 1. Panel (a) shows simulated data with deterministic location by group membership. Panel (b) shows simulated data where half of the observations are located in a deterministic fashion and the remainder are randomly located.The structure from Panel (b) is revisited in simulated environments.

Table 1
Homogeneous spatial dependence estimation results $\rho = .7$.

Parameter	True values	SAR-M	Std Dev	L95★	U95★	OLS-M	Std Dev	L95★	U95★
β_{11}	−0.5000	−0.5091	0.0174	−0.5435	−0.4756	−0.7650	0.0468	−0.8543	−0.6731
β_{12}	−0.7500	−0.7506	0.0158	−0.7813	−0.7188	−1.0031	0.0473	−1.0966	−0.9074
β_{21}	0.5000	0.5091	0.0175	0.4745	0.5423	0.6163	0.0754	0.4686	0.7661
β_{22}	0.8000	0.8014	0.0200	0.7596	0.8406	0.9311	0.0798	0.7661	1.0846
β_{31}	−1.0000	−0.9845	0.0253	−1.0327	−0.9313	−1.0580	0.1364	−1.3267	−0.7785
β_{32}	1.2000	1.1580	0.0249	1.1072	1.2026	1.2773	0.1291	1.0166	1.5172
ρ	0.7000	0.7097	0.0084	0.6934	0.7259	–	–	–	–
σ_1^2	1.0000	1.0481	0.0999	0.8972	1.2838	6.5587	0.5700	5.5326	7.7521
σ_2^2	0.7500	0.9236	0.1436	0.721	1.2845	8.2107	0.7772	6.7451	9.7651
σ_3^2	0.5000	0.8295	0.2152	0.5338	1.3564	9.7612	1.3678	7.1569	12.5825
π_1	0.4500	0.4512	0.0136	0.4241	0.4785	0.4456	0.0195	0.4074	0.4833
π_2	0.3500	0.3491	0.0137	0.3223	0.3761	0.3443	0.0280	0.2905	0.3987
π_3	0.2000	0.1997	0.0131	0.1737	0.2257	0.2102	0.0290	0.1538	0.2672
r^2	0.9326								
R^2		0.9381				0.6904			

Note: ★ L95 and U95 represent the lower and upper bounds of the 95% highest posterior density.

means reflective of the true value but the 95% credible interval suggests each result is significantly different from zero and that the variation around that mean is relatively low. The calculated R^2 is close to the established r^2 indicating that the model is in fact capturing the vast majority of the signal. It is important to note that the values for both R^2 and r^2 are the average across the simulations. Fig. 4 shows diagnostic plots of the draws for each parameter with no evidence of label switching or other convergence issues. OLS results are omitted but suffer from the same bias as shown previously.

It should be noted that here the direct and indirect effects, representing own-partial and cross-partial derivatives respectively have been omitted from the table. This is a purposeful omission in that these effects hold little meaning in a simulation environment though later, in the empirical results, the scalar summaries for these effects will be provided. Furthermore, while these values have been calculated within the confines of the empirical example there is significant discussion to be had regarding the interpretation of such derivatives in the mixture model framework, discussion that will be saved for a later date.

Referring back to Table 2 it is apparent that the homogeneous SAR estimates are quite different from the true values (TV) and should be seen as unreliable. The estimate for ψ in this model is quite a bit larger than the weighted average of the scalar parameters in the heterogeneous model (.09). In addition, the estimates for β not only are different from the group weighted average but are significantly different from zero, supporting the assertion originally made by Anselin and Arribas-Bel (2013) that spurious results are a common result of model misspecification between spatial heterogeneity and spatial dependence.

Tables 3–5 illustrate how the model performs as the noise is increased in the DGP. Note that as the level of signal decreases the estimates from the SAR-M model become more uncertain with larger standard deviations and wider credible intervals until, in the lowest signal case, the estimates are no longer statistically distinguishable from zero. The SAR on the other hand, behaves consistently regardless of the noise level. Standard deviations increase slightly for each parameter as the level of noise increases though one would continue to reject the null hypothesis of zero in each case. This is related to the well-known problem of bias-variance trade-off. Using multiple mixtures fits the DGP accurately, at the cost of extra parameters. With more noise, it is harder to estimate them precisely leading to higher variance in the estimates.

Table 2
High signal simulations.

	TV	SAR-M	Std Dev	L95*	U95*		SAR	Std Dev	L95*	U95*
β_{11}	−0.5000	−0.4992	0.0234	−0.5445	−0.4527	β_1	−0.2597	0.0508	−0.3576	−0.1579
β_{12}	−0.7500	−0.7516	0.0220	−0.7944	−0.7085	β_2	0.2529	0.0710	0.1070	0.3880
β_{21}	0.5000	0.4944	0.0235	0.4483	0.5399	–	–	–	–	–
β_{22}	0.8000	0.7957	0.0244	0.7467	0.8422	–	–	–	–	–
β_{31}	−1.0000	−0.9787	0.0314	−1.0373	−0.9133	–	–	–	–	–
β_{32}	1.2000	1.1804	0.0320	1.1145	1.2400	–	–	–	–	–
σ_1^2	1.0000	1.0576	0.1236	0.8650	1.3508	σ^2	7.5598	0.4013	6.7954	8.3697
σ_2^2	0.7500	0.8841	0.1593	0.6451	1.2685	–	–	–	–	–
σ_2^2	0.5000	0.7767	0.1823	0.4931	1.2053	–	–	–	–	–
ρ_1	−0.3000	−0.2767	0.0404	−0.3576	−0.1992	ρ	0.2986	0.0532	0.1930	0.4010
ρ_2	0.3000	0.2996	0.0328	0.2350	0.3640	–	–	–	–	–
ρ_3	0.6000	0.5902	0.0434	0.5046	0.6756	–	–	–	–	–
π_1	0.4500	0.4346	0.0186	0.3980	0.4703	–	–	–	–	–
π_2	0.3500	0.3511	0.0196	0.3140	0.3902	–	–	–	–	–
π_3	0.2000	0.2143	0.0177	0.1797	0.2492	–	–	–	–	–
R^2	0.8862	–	–	–	–	–	–	–	–	–
r^2	–	0.8598	–	–	–	–	0.0356	–	–	–

Note: * L95 and U95 represent the lower and upper bounds of the 95% highest posterior density.

Table 3
Simulations: medium-high signal.

	TV	SAR-M	Std Dev	L95*	U95*		SAR	Std Dev	L95*	U95*
β_{11}	−0.5000	−0.4800	0.0690	−0.6222	−0.3508	β_1	−0.2479	0.0529	−0.3482	−0.1402
β_{12}	−0.7500	−0.7831	0.0675	−0.9121	−0.6439	β_2	0.2672	0.0678	0.1326	0.4001
β_{21}	0.5000	0.4836	0.0777	0.3295	0.6364	–	–	–	–	–
β_{22}	0.8000	0.7708	0.0756	0.6173	0.9140	–	–	–	–	–
β_{31}	−1.0000	−0.8905	0.0974	−1.0841	−0.7084	–	–	–	–	–
β_{32}	1.2000	1.0342	0.1128	0.7942	1.2342	–	–	–	–	–
σ_1^2	6.0000	5.8623	0.6426	4.7283	7.2407	σ^2	10.2862	0.4072	9.5134	11.1418
σ_2^2	4.5000	5.0891	0.7249	3.7642	6.6358	–	–	–	–	–
σ_2^2	3.0000	4.6521	0.7322	3.3034	6.1890	–	–	–	–	–
ρ_1	−0.3000	−0.2034	0.0863	−0.3748	−0.0227	ρ	0.2464	0.0330	0.1810	0.3100
ρ_2	0.3000	0.3529	0.0642	0.2315	0.4753	–	–	–	–	–
ρ_3	0.6000	0.6896	0.0866	0.5043	0.8483	–	–	–	–	–
π_1	0.4500	0.3944	0.0274	0.3420	0.4493	–	–	–	–	–
π_2	0.3500	0.3510	0.0326	0.2845	0.4120	–	–	–	–	–
π_3	0.2000	0.2546	0.0359	0.1871	0.3270	–	–	–	–	–
R^2	0.6699	–	–	–	–	–	–	–	–	–
r^2	–	0.6177	–	–	–	–	0.0257	–	–	–

Note: * L95 and U95 represent the lower and upper bounds of the 95% highest posterior density.

Table 4
Simulations: medium-low signal.

	TV	SAR-M	Std Dev	L95*	U95*		SAR	Std Dev	L95*	U95*
β_{11}	−0.5000	−0.4304	0.2310	−0.7936	0.2805	β_1	−0.2487	0.0610	−0.3653	−0.1257
β_{12}	−0.7500	−0.6817	0.1963	−1.0490	−0.2929	β_2	0.2343	0.0726	0.0874	0.3762
β_{21}	0.5000	0.3585	0.2883	−0.5719	0.7602	–	–	–	–	–
β_{22}	0.8000	0.6386	0.2683	−0.0608	1.0598	–	–	–	–	–
β_{31}	−1.0000	−0.8659	0.1661	−1.1562	−0.5494	–	–	–	–	–
β_{32}	1.2000	0.8543	0.2208	0.3973	1.2553	–	–	–	–	–
σ_1^2	15.0000	15.1365	2.2533	11.3241	20.1626	σ^2	14.5948	0.5877	13.5026	15.7993
σ_2^2	12.0000	14.0234	2.0049	10.5601	18.3847	–	–	–	–	–
σ_2^2	9.0000	12.9411	1.8492	9.7528	16.9530	–	–	–	–	–
ρ_1	−0.3000	−0.2299	0.1537	−0.5454	0.0520	ρ	0.2715	0.0331	0.2080	0.3350
ρ_2	0.3000	0.3786	0.1099	0.1259	0.5729	–	–	–	–	–
ρ_3	0.6000	0.6523	0.1322	0.3863	0.8833	–	–	–	–	–
π_1	0.4500	0.3754	0.0583	0.2573	0.4945	–	–	–	–	–
π_2	0.3500	0.3346	0.0562	0.2206	0.4508	–	–	–	–	–
π_3	0.2000	0.2899	0.0633	0.1671	0.4201	–	–	–	–	–
R^2	0.4354	–	–	–	–	–	–	–	–	–
r^2	–	0.3440	–	–	–	–	0.0187		–	–

Note: * L95 and U95 represent the lower and upper bounds of the 95% highest posterior density.

Table 5
Simulations: low signal.

	TV	SAR-M	Std Dev	L95*	U95*		SAR	Std Dev	L95*	U95*
β_{11}	−0.5000	−0.2738	0.4119	−0.9991	0.6214	β_1	−2.1660	0.0790	−0.3719	−0.0599
β_{12}	−0.7500	−0.5556	0.3238	−1.1078	0.3197	β_2	0.2379	0.0824	0.0795	0.3988
β_{21}	0.5000	0.0996	0.4525	−0.9023	0.8223	–	–	–	–	–
β_{22}	0.8000	0.5081	0.4290	−0.5531	1.1826	–	–	–	–	–
β_{31}	−1.0000	−0.8091	0.2767	−1.2488	−0.2097	–	–	–	–	–
β_{32}	1.2000	0.8048	0.3181	0.0264	1.3029	–	–	–	–	–
σ_1^2	24.0000	25.1783	3.9301	18.7115	34.1922	σ^2	22.1068	0.8985	20.4180	23.9257
σ_2^2	21.0000	23.9616	3.5976	17.9829	32.2192	–	–	–	–	–
σ_2^2	18.0000	22.5176	3.1707	17.2239	29.7827	–	–	–	–	–
ρ_1	−0.3000	−0.2178	0.1841	−0.5533	0.1167	ρ	0.1866	0.0341	0.1190	0.2540
ρ_2	0.3000	0.4061	0.1575	0.0453	0.6770	–	–	–	–	–
ρ_3	0.6000	0.5713	0.1217	0.3408	0.8043	–	–	–	–	–
π_1	0.4500	0.3769	0.0822	0.2116	0.5377	–	–	–	–	–
π_2	0.3500	0.3282	0.0716	0.1817	0.4744	–	–	–	–	–
π_3	0.2000	0.2949	0.0695	0.1583	0.4400	–	–	–	–	–
R^2	0.3115	–	–	–	–	–	–	–	–	–
r^2	–	0.1772	–	–	–	–	0.0113		–	–

Note: * L95 and U95 represent the lower and upper bounds of the 95% highest posterior density.

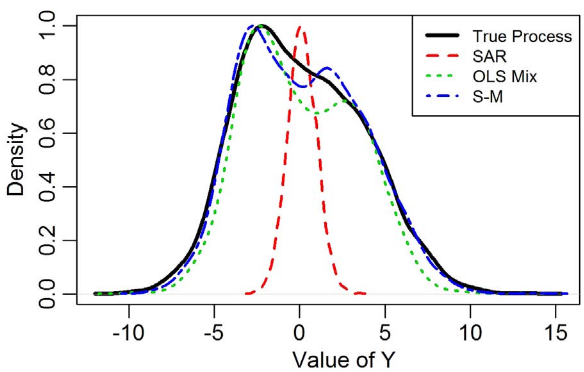


Fig. 2. The solid line in Figure (2) specifically illustrates the signal in the simulated data. Each model selection is compared to the true signal via the estimate $\hat{y}$. This figure clearly shows that homogeneous SAR models are unable to adequately capture the available signal in the simulated data.

Fig. 2 provides better insight as to why the SAR estimates have consistently low variance but large bias. This figure plots the fitted values for each model type against the true process of y. The SAR model in this case captures very little of the process due to clear misspecification and captures roughly the same level of information regardless of the R^2 used. In effect, the SAR is capturing a small amount of signal and can continue to capture that same level throughout the simulation environments. It should be stressed that the results here are expected for the SAR model as both spatial heterogeneity and spatial dependence are present in the DGP by construction. The results are misleading and spurious relative to the known values.

5. Empirical application

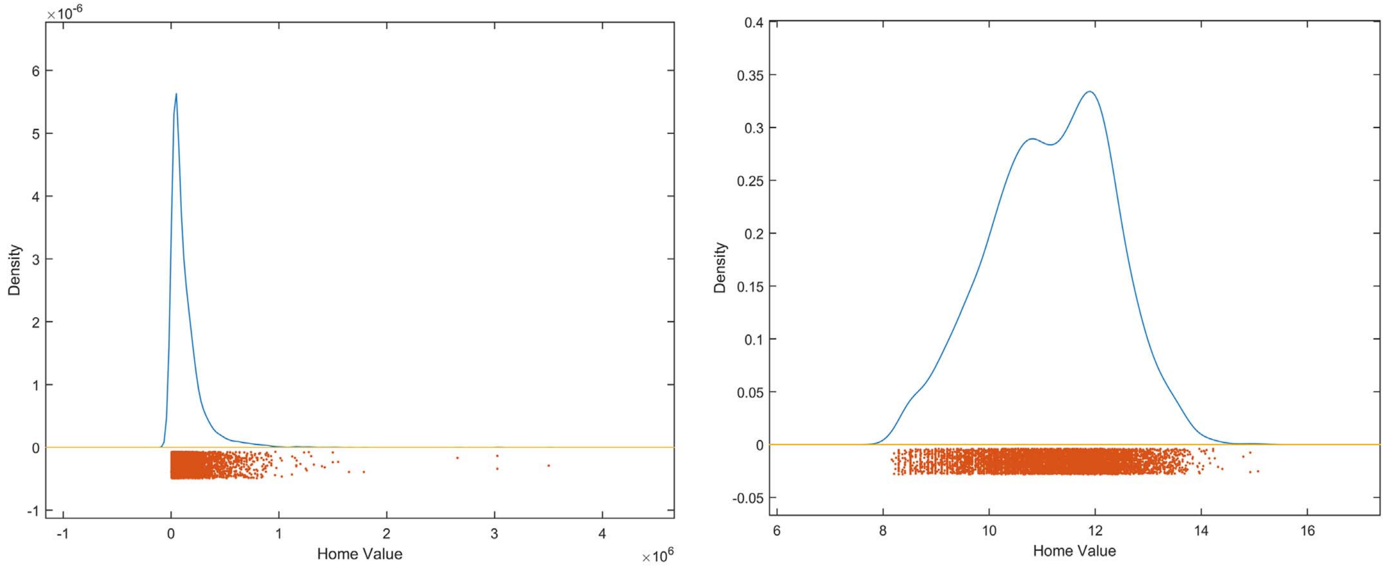
Lately there have been a number of empirical studies which utilize spatial econometric methods to estimate hedonic pricing in real estate. It is within this domain that the SAR-M model will be illustrated. By reviewing the housing literature using the Social Science Citation Index, (Kuminoff et al., 2010) underline that spatial autocorrelation is controlled by fixed effects for more than 50 percent of hedonic price analyses. It is often assumed to be a preferable process in the face of spatially correlated omitted variables. Directly refuting this assertion Anselin and Arribas-Bel (2013) clearly illustrate that fixed effects models do not properly account for spatial dependence in most cases. The illustration presented here will show that both spatial dependence and heterogeneity can be properly estimated with mixture models. As underlined in the Monte Carlo experiments, spatial regression models not robust to mixture distributions may provide spurious results and artificially create spatial dependence. This illustration will confirm that ignoring the SAR-M model in favor of standard spatial based models will produce a biased estimate for ρ generally in an upward fashion. Ignoring the spatial dependence however produces biased parameter estimates even in the mixture case.

The housing dataset used here was obtained from the Hamilton County Auditor. Sale price for the year 2011 has been merged with an extensive database of residential single-family properties in the City of Cincinnati. This database contains information pertaining to the

Table 6
Data statistics and definitions.

Variable	Mean	Std. Dev.	Min	Max	Description
Price*	127.00	167.00	3.5	3500.00	Sales price in 2011 of single-family residential properties (Hamilton County Auditor)
DistCBD	10.02	0.54	7.52	10.92	Distance to Downtown Cincinnati (calculated in ArcView)
Sqftland	5431.66	2.12	475.61	433,653.00	Size of parcel in square feet (Hamilton County Auditor)
Sqftbuilt	992.28	1.36	403.99	6438.17	Size of floor space (Hamilton County Auditor)
Numbdrooms	3.07	1.09	0.00	20.00	Total number of bedrooms (Hamilton County Auditor)
Age	86.23	30.15	0.00	198.00	Age of the house in years (Hamilton County Auditor)
Full Baths	1.59	0.75	0	9	Number of full bathrooms
Central Air	0.63	0.48	0	1	Dummy indicating if home has central air
Age65	11.51	4.68	0.00	28.10	Percent population 65 years and older (2010 Census)
Renter	51.95	17.32	3.22	97.64	Percentage of renters (2010 Census)
HSdegree	84.69	10.30	58.50	100.00	Percent population with High School degree (2010 Census)

*: In 1000's of dollars.



(a) Density: Home Price (b) Density: Log of Home Price

Fig. 3. Panel (a) shows distribution of the home prices prior to transformation. As expected this distribution is heavily skewed with a long right tail. Panel (b) shows the post transformation home prices. Clear bi-modality and thick tails are visible.

Table 7
SAR-M estimates.

	Group 1				Group 2			
	Mean	Std. Dev	L95⋆	U95⋆	Mean	Std. Dev	L95⋆	U95⋆
Intercept	4.6838	0.5922	3.4710	5.8021	7.4447	0.9028	5.6750	9.2276
distCBD	−0.4226	0.0368	−0.4946	−0.3503	−0.2872	0.0594	−0.4024	−0.1711
logsqftland	0.1664	0.0367	0.0962	0.2431	−0.0477	0.0453	−0.1337	0.0447
logsqftbuild	0.2375	0.0647	0.1144	0.3674	0.2508	0.0992	0.0517	0.4386
Numbdrms	0.0728	0.0144	0.0455	0.1022	0.0134	0.0205	−0.0271	0.0534
House Age	−0.0018	0.0006	−0.0030	−0.0007	−0.0141	0.0012	−0.0164	−0.0117
Full Baths	0.1420	0.0229	0.0970	0.1867	0.1115	0.0342	0.0444	0.1782
Central Air	0.5030	0.0365	0.4329	0.5766	0.5558	0.0544	0.4490	0.6602
Age 65	0.0035	0.0028	−0.0023	0.0090	−0.0042	0.0055	−0.0152	0.0062
Renter	−0.0007	0.0010	−0.0028	0.0014	−0.0021	0.0018	−0.0056	0.0016
Hsdegree	0.0654	0.0028	0.0598	0.0708	0.0459	0.0039	0.0378	0.0533
σ^2	0.1867	0.0158	0.1572	0.2189	0.4491	0.0306	0.3931	0.5124
π	0.4283	0.0217	0.3846	0.4693	0.2692	0.0241	0.2234	0.3169
	Group 3				Group 4			
	Mean	Std. Dev	L95⋆	U95⋆	Mean	Std. Dev	L95⋆	U95⋆
Intercept	6.8459	0.6025	5.6820	8.0549	12.9903	2.9694	6.7780	18.6099
distCBD	−0.4157	0.0409	−0.4959	−0.3354	−0.3764	0.2596	−0.8569	0.1797
logsqftland	0.0644	0.0267	0.0179	0.1227	−0.0410	0.1502	−0.2649	0.3596
logsqftbuild	0.3737	0.0681	0.2401	0.5089	−0.0177	0.3521	−0.7635	0.6609
Numbdrms	0.0760	0.0198	0.0353	0.1144	0.0462	0.0954	−0.1424	0.2349
House Age	0.0010	0.0006	−0.0001	0.0022	0.0033	0.0027	−0.0018	0.0087
Full Baths	0.1967	0.0306	0.1357	0.2567	0.4521	0.1760	0.0917	0.7807
Central Air	0.1770	0.0436	0.0864	0.2592	−0.6142	0.1927	−1.0083	−0.2451
Age 65	0.0110	0.0033	0.0046	0.0175	0.0229	0.0195	−0.0161	0.0606
Renter	−0.0017	0.0012	−0.0041	0.0008	0.0010	0.0070	−0.0115	0.0164
Hsdegree	0.0423	0.0023	0.0374	0.0465	0.0043	0.0091	−0.0143	0.0219
σ^2	0.1516	0.0168	0.1219	0.1877	0.7072	0.1453	0.4731	1.0446
π	0.2532	0.0237	0.2091	0.3000	0.0493	0.0079	0.0349	0.0654
ρ	0.1343	0.0163	0.1120	0.1747				
Num. Obs.	7574							
DIC	−16,305							

Note: ⋆ L95 and U95 represent the lower and upper bounds of the 95% highest posterior density.

structural characteristics of each home. The data set is comprised of 7574 residential properties sold in 2011. In addition to the sales information and structural characteristics, a number of neighborhood attributes have been included using Census data to capture the impact of local amenities on real estate prices. Neighborhoods are delineated by the 116 census tracts within the city limits and a log-linear functional form is used to control for large variation in sale price. Variables along with some descriptive statistics are presented in Table 6.

The distribution of observed sale prices is shown to have a multi-

Table 8
SAR-M: partial effects estimates.

		Group 1			Group 2			Group 3			Group 4		
		Mean	L95*	U95*	Mean	L95*	U95*	Mean	L95*	U95*	Mean	L95*	U95*
	distCBD	−0.4882	−0.6014	−0.3760	−0.3319	−0.5108	−0.1529	−0.4802	−0.6019	−0.3597	−0.4345	−1.2175	0.4818
	logsqftland	0.1923	0.0893	0.3141	−0.0550	−0.1803	0.0838	0.0744	0.0054	0.1807	−0.0472	−0.3909	0.6167
	logsqftbuild	0.2744	0.0834	0.4728	0.2898	−0.0172	0.5816	0.4317	0.2199	0.6305	−0.0209	−1.2386	1.0199
	Numbdrms	0.0842	0.0419	0.1274	0.0156	−0.0471	0.0743	0.0879	0.0258	0.1449	0.0534	−0.2318	0.3463
Total	House Age	−0.0021	−0.0039	−0.0003	−0.0163	−0.0197	−0.0130	0.0012	−0.0005	0.0031	0.0038	−0.0038	0.0127
	Full Baths	0.1641	0.0932	0.2312	0.1288	0.0235	0.2287	0.2272	0.1372	0.3209	0.5221	−0.0373	1.0284
	Central Air	0.5811	0.4726	0.6897	0.6423	0.4702	0.8104	0.2043	0.0550	0.3256	−0.7099	−1.3416	−0.1583
	Age 65	0.0040	−0.0048	0.0122	−0.0049	−0.0216	0.0110	0.0127	0.0030	0.0224	0.0264	−0.0331	0.0859
	Renter	−0.0008	−0.0041	0.0023	−0.0024	−0.0079	0.0033	−0.0019	−0.0059	0.0021	0.0012	−0.0178	0.0259
	Hsdegree	0.0756	0.0691	0.0833	0.0530	0.0403	0.0646	0.0489	0.0423	0.0545	0.0050	−0.0246	0.0327
	distCBD	−0.0644	−0.0968	−0.0453	−0.0438	−0.0792	−0.0190	−0.0632	−0.0937	−0.0433	−0.0570	−0.1736	0.0687
	logsqftland	0.0253	0.0110	0.0445	−0.0072	−0.0247	0.0117	0.0098	0.0007	0.0246	−0.0061	−0.0573	0.0835
	logsqftbuild	0.0362	0.0106	0.0713	0.0383	−0.0022	0.0864	0.0569	0.0286	0.0937	−0.0032	−0.1715	0.1362
	Numbdrms	0.0111	0.0053	0.0193	0.0021	−0.0060	0.0110	0.0116	0.0032	0.0217	0.0071	−0.0314	0.0479
Indirect	House Age	−0.0003	−0.0006	0.0000	−0.0021	−0.0031	−0.0015	0.0002	−0.0001	0.0005	0.0005	−0.0005	0.0018
	Full Baths	0.0216	0.0114	0.0366	0.0169	0.0030	0.0326	0.0300	0.0169	0.0493	0.0686	−0.0054	0.1483
	Central Air	0.0766	0.0549	0.1095	0.0848	0.0567	0.1329	0.0268	0.0073	0.0476	−0.0938	−0.2020	−0.0203
	Age 65	0.0005	−0.0006	0.0018	−0.0006	−0.0030	0.0015	0.0017	0.0004	0.0032	0.0035	−0.0046	0.0121
	Renter	−0.0001	−0.0006	0.0003	−0.0003	−0.0011	0.0004	−0.0002	−0.0008	0.0003	0.0002	−0.0025	0.0038
	Hsdegree	0.0099	0.0080	0.0133	0.0070	0.0047	0.0108	0.0064	0.0051	0.0087	0.0006	−0.0036	0.0043
	distCBD	−0.4238	−0.5223	−0.3266	−0.2881	−0.4397	−0.1342	−0.4169	−0.5209	−0.3102	−0.3775	−1.0543	0.4165
	logsqftland	0.1670	0.0776	0.2755	−0.0478	−0.1574	0.0731	0.0646	0.0047	0.1575	−0.0411	−0.3386	0.5408
	logsqftbuild	0.2382	0.0709	0.4128	0.2516	−0.0150	0.5030	0.3748	0.1915	0.5468	−0.0177	−1.0633	0.8969
	Numbdrms	0.0731	0.0365	0.1104	0.0135	−0.0412	0.0647	0.0763	0.0224	0.1249	0.0463	−0.2009	0.3006
Direct	House Age	−0.0018	−0.0034	−0.0003	−0.0141	−0.0172	−0.0112	0.0010	−0.0005	0.0026	0.0033	−0.0034	0.0109
	Full Baths	0.1424	0.0807	0.2009	0.1118	0.0205	0.1980	0.1973	0.1181	0.2780	0.4535	−0.0328	0.8928
	Central Air	0.5045	0.4135	0.6005	0.5575	0.4105	0.6981	0.1775	0.0477	0.2838	−0.6161	−1.1584	−0.1406
	Age 65	0.0035	−0.0042	0.0107	−0.0042	−0.0189	0.0096	0.0111	0.0026	0.0195	0.0229	−0.0287	0.0746
	Renter	−0.0007	−0.0036	0.0020	−0.0021	−0.0069	0.0028	−0.0017	−0.0050	0.0018	0.0010	−0.0154	0.0221
	Hsdegree	0.0656	0.0579	0.0734	0.0460	0.0352	0.0556	0.0425	0.0356	0.0479	0.0043	−0.0212	0.0288

Note: * L95 and U95 represent the lower and upper bounds of the 95% highest posterior density.

Table 9
SAR results.

	Mean	Std. Dev	L95*	U95*
Intercept	2.3329	0.2637	1.8043	2.8377
distCBD	−0.218	0.0181	−0.2538	−0.1827
logsqftland	0.0739	0.0118	0.0511	0.0967
logsqftbuild	0.2288	0.0284	0.1721	0.2827
Numbdrms	0.0392	0.0084	0.0228	0.0555
House Age	−0.0021	0.0003	−0.0027	−0.0016
Full Baths	0.1025	0.0123	0.0782	0.1254
Central Air	0.3399	0.0175	0.3038	0.3727
Age 65	−0.0001	0.0016	−0.0032	0.0032
Renter	−0.0007	0.0005	−0.0017	0.0003
Hsdegree	0.0281	0.001	0.0261	0.0301
σ^2	0.2607	0.0055	0.2506	0.2719
ρ	0.5518	0.0099	0.533	0.571

Note: * L95 and U95 represent the lower and upper bounds of the 95% highest posterior density.

modal structure, as plotted in Fig. 3, which provides at least weak evidence of a possible mixture structure. Even though plot (3a) does not reveal a multi-modal distribution directly it does indicate one which is skewed with a large right tail. The log of that same data (plot (3b)) exhibits some level of multi-modality with two distinct modes readily visible. Fitting a mixture of normals to the observed housing price will accommodate both data structures though this example makes clear the dangers of taking an inferential leap and equating visual modes to the number of component mixtures under a cursory examination of the data. Even if sale prices were to be tightly clustered, the within-cluster distribution of observed prices may be non-normal and provide evidence that a mixture of distributions is a more adequate representation of the true process. Heterogeneity is thought to be caused not by the unobserved variation in household preferences but mainly by housing quality that is unobserved by the econometrician (Epple et al., 2014).

Tables 7 and 9 cover the results of the SAR-M and SAR specifications respectively. The SAR-M is presented with four component distributions, a decision made by comparing Deviance Information Criterion (DIC) values along with diagnostic plots for several competing specifications. Alternative specifications included models with up to six component distributions. While DIC indicated a model with five components may ultimately provide a better fit, the four component mixture model was selected. This is the result of examining diagnostic plots, parameter estimates, and 95% highest-posterior-density (HPD) intervals for both specifications.[6] Under a five component model the parameter estimates for the fifth group were not significantly different from zero, the 95% HPDs for the remaining groups increased relative to the four component case, and there was no statistical change in the estimate of ρ. As noted by Frühwirth-Schnatter (2004), the marginal likelihood is generally a more accurate and applicable model selection tool when evaluating mixture models provided it is possible to derive an explicit form from which to make comparisons.

The SAR specification indicates that a majority of the variables are significantly different than zero with well-defined 95% credible intervals. To properly interpret the SAR-M model, partial derivatives of the reduced form (23) have to be calculated. Conditional on the indicator matrix z, partial derivatives are equal to

$$\frac{\delta y}{\delta X'_k} = (I_N - \widetilde{\Psi} W)^{-1} \operatorname{diag}(z\beta_k), \tag{28}$$

where β_k is a G-dimensional vector of parameters. LeSage and Pace (2009) define the direct effect as the average of the diagonal elements

[6] Estimation results for all number of mixture components along with AIC, BIC, DIC and diagnostic plots are available upon request.

Table 10
SAR: partial effects.

	Direct			Indirect			Total		
	Mean	L95*	U95*	Mean	L95*	U95*	Mean	L95*	U95*
distCBD	−0.2349	−0.2737	−0.1969	−0.2517	−0.2972	−0.2064	−0.4866	−0.5669	−0.4060
logsqftland	0.0797	0.0550	0.1041	0.0854	0.0586	0.1125	0.1650	0.1147	0.2166
logsqftbuild	0.2466	0.1849	0.3049	0.2643	0.1948	0.3330	0.5109	0.3810	0.6352
Numbdrms	0.0423	0.0244	0.0598	0.0453	0.0252	0.0645	0.0875	0.0500	0.1238
House Age	−0.0023	−0.0029	−0.0018	−0.0025	−0.0031	−0.0019	−0.0047	−0.0060	−0.0036
Full Baths	0.1104	0.0842	0.1353	0.1183	0.0889	0.1477	0.2287	0.1736	0.2815
Central Air	0.3662	0.3274	0.4019	0.3924	0.3439	0.4422	0.7586	0.6734	0.8422
Age 65	−0.0001	−0.0035	0.0035	−0.0001	−0.0037	0.0038	−0.0002	−0.0072	0.0073
Renter	−0.0007	−0.0018	0.0004	−0.0008	−0.0020	0.0004	−0.0015	−0.0038	0.0007
Hsdegree	0.0303	0.0281	0.0325	0.0325	0.0293	0.0359	0.0628	0.0575	0.0682

Note: * L95 and U95 represent the lower and upper bounds of the 95% highest posterior density.

of the matrix of partial derivatives (28), and the indirect effect as the average of the cumulative off-diagonal elements from each row. Since, in the case of a heterogeneous model, there is the potential for considerable variation across parameter estimates (including ρ) these summaries are averaged over observations within the same sub-group. Table 8 provides scalar summaries of the own-partial (direct) and cross-partial (indirect) derivatives outlined by group from the SAR-M specification. It should be noted that additional direction on the interpretation of these effects is forthcoming (Cornwall, 2016) (Table 9). This paper will further decompose the outlined partial effects into group level spill-in and spill-out effects (see LeSage and Chih, 2016). The scalar summaries for the SAR specification are outlined in Table 10.

Under the SAR specification, the largest positive impacts are produced by the direct effects of home size (logsqftbuild) and the presence of central air. Also important are the number of full bathrooms (Full Baths) though to a lesser extent when compared to the aforementioned attributes. The largest negative impact on home values is measured distance to the central business district highlighting once again that location in real estate is incredibly important.

The results of the SAR-M model tell a similar but richer story. The SAR model failed to reject the null that $\beta_{Age65} = 0$ while the SAR-M rejects this null for group three. Increasing the proportion of residents aged 65 or older by 1% increases the home value by approximately 1.27% for this specific group. One possible explanation for such a relationship is that older residents tend to value externalities created by similar neighbors (e.g. reduced noise pollution) and as a result are willing to pay some small premium to secure such neighbors.

Perhaps more interesting is the parameter estimates of Central Air, the presence of which generally leads to increased home values. The SAR-M supports this for the majority of the sample however for homes in group four the presence of Central Air has a large negative effect on home value. Homes found in group four tend to be older than the average home in each of the other groups, averaging nearly 90 years, indicating some of these homes may be designated as historic sites. Hamilton County has separate rules which govern historic homes and the improvements that can be made to such properties.[7] This is one possible explanation for the negative effect Central Air has on home value (for group 4); adding it may cause a loss in historical designation and by extension value.

Consider that for some of the variables in the SAR-M specification the group parameters are not only similar in mean but overlap heavily with respect to their highest posterior density. This homogeneity across groups indicates that there may be some larger regional preference set influencing home prices based on these attributes. Distance to the central business district appears to have a relatively homogeneous effect on home prices (with the exception of those found in group four) as does the number of full bathrooms. Note that these homogeneous variables tend to have directionality consistent previous research. For instance, a lower number of bathrooms tends to lead to a lower home value *ceteris paribus*.

Different intercepts across the groups within the SAR-M results tell us that the homes included in group four benefit more from unobservable traits than the other three groups. This supports the idea that, within this data set, there is some level of spatial heterogeneity which would be identifiable if appropriate borders could be constructed in a spatial fixed effects specification. Of course the heterogeneous nature of the coefficients and resulting partial derivatives would be missed in such a model. (Fig. 4).

Fig. 5 outlines the classification probabilities as defined in (21). Here it can be seen that there is in fact some clustering that is creating higher spatial intensity within areas, again leading to the belief that spatial heterogeneity is present within the data. Homes in the eastern part of Hamilton County are more likely to be in group one though pockets of similar homes appear all over the map. The same can be said for group four which had the most drastic difference in intercept though many of these homes seem to be most probably located in the southwest and central portion of the county with pockets existing throughout the remainder of the space.

It should be pointed out that only one value of ρ is outlined for the SAR-M model. Testing showed that each group had statistically indistinguishable ρ values for each group and the variance of other parameters was reduced by constraining the model to a single ρ parameter over the full data set.[8] Since the SAR-M model nests a number of specification cases it is important to consider the dynamics of the market being modeled. In this case, it seemed appropriate that the housing market would have similar levels of spatial dependency across groups in such a small geographic area and as a result the constrained model was chosen.

6. Conclusions

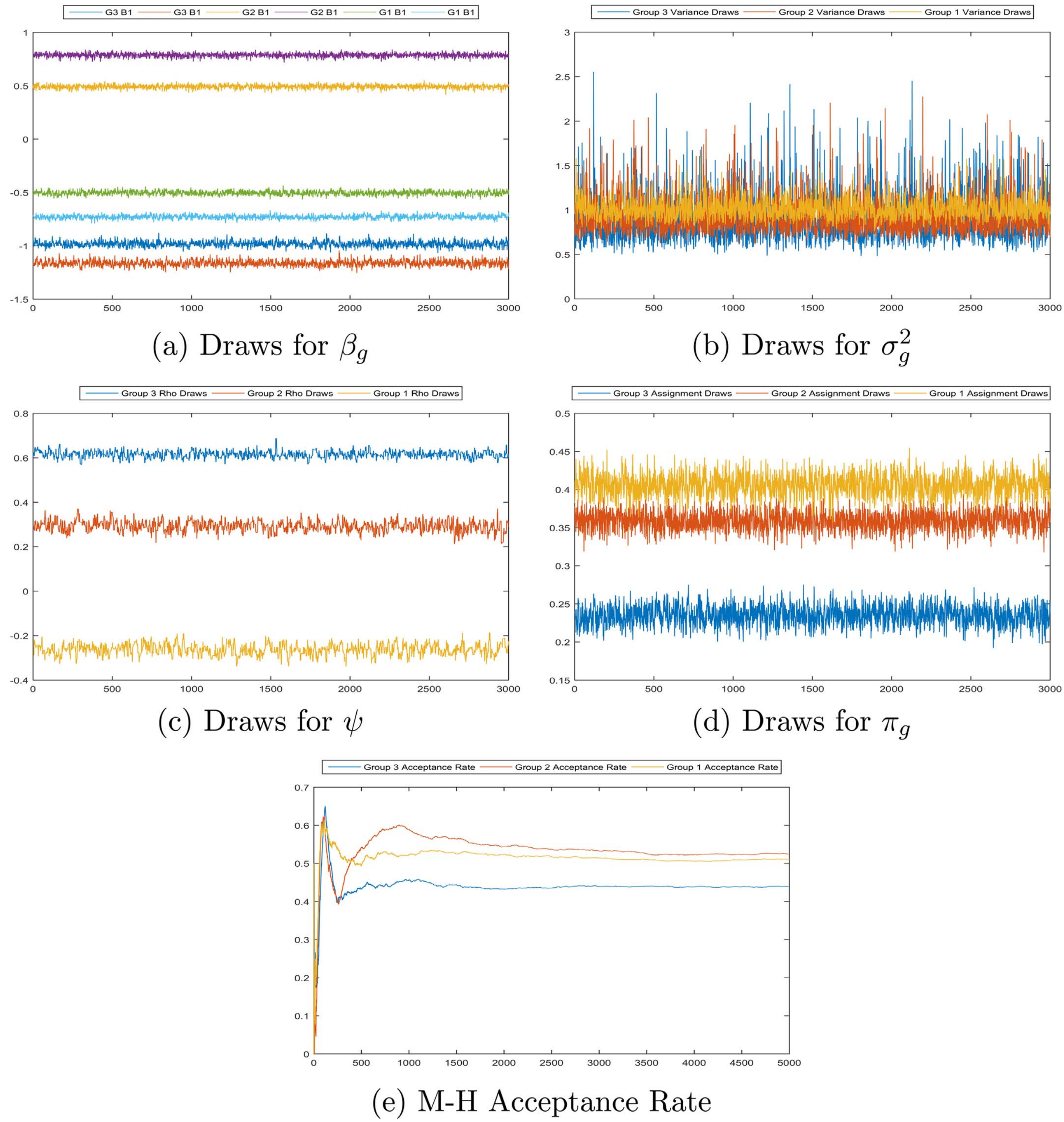
This paper develops a method whereby both spatial heterogeneity and dependence are controlled for within the same construct. Previously, these two issues were handled separately and have been shown to provide biased or spurious results if the correct specification was not chosen. The method presented here has a number of advantages. First, it is easily implemented within the existing frame-

[7] See http://www.cincinnati-oh.gov/buildings/historic-conservation/ for one such example of local regulations governing historic properties.

[8] Each had a mean of approximately .135 with nearly identical credible intervals. Odds against $\rho_g = .135$ are 1.6374 to 1, 1.2560 to 1, 1.0409 to 1, and 1.0323 to 1 respectively where 0.135 is the approximate value found in the constrained SAR-M model. These results clearly indicate that the scalar value is statistically indistinguishable between the groups and led to the constrained homogeneous SAR-M model choice.

work of mixture models in econometric literature. This allows the researcher to lean on a robust line of research outlining both inference and selection criteria for these models. Second, the incorporation of the spatial framework is quite flexible and while this particular exposition focused on the SAR specification the SAR-M model can be extended to both the Spatial Durbin Error (SDEM) and Spatial Durbin Model (SDM) structures.

Spatial mixture models have a number of potential applications. A latent indicator variable is used to capture population heterogeneity thereby allowing researchers to evaluate data that exhibits both spatial heterogeneity and dependence. Even if the population heterogeneity is well-thought-out *a priori* and the underlying causal process seems to be well-delineated, it is unlikely that the researcher will be able to identify all of the important predictors. A mixture structure can be used to explore a data set for evidence of clusters characterized by differential effects and provide a more robust interpretation of the model given data.



(a) Draws for β_g

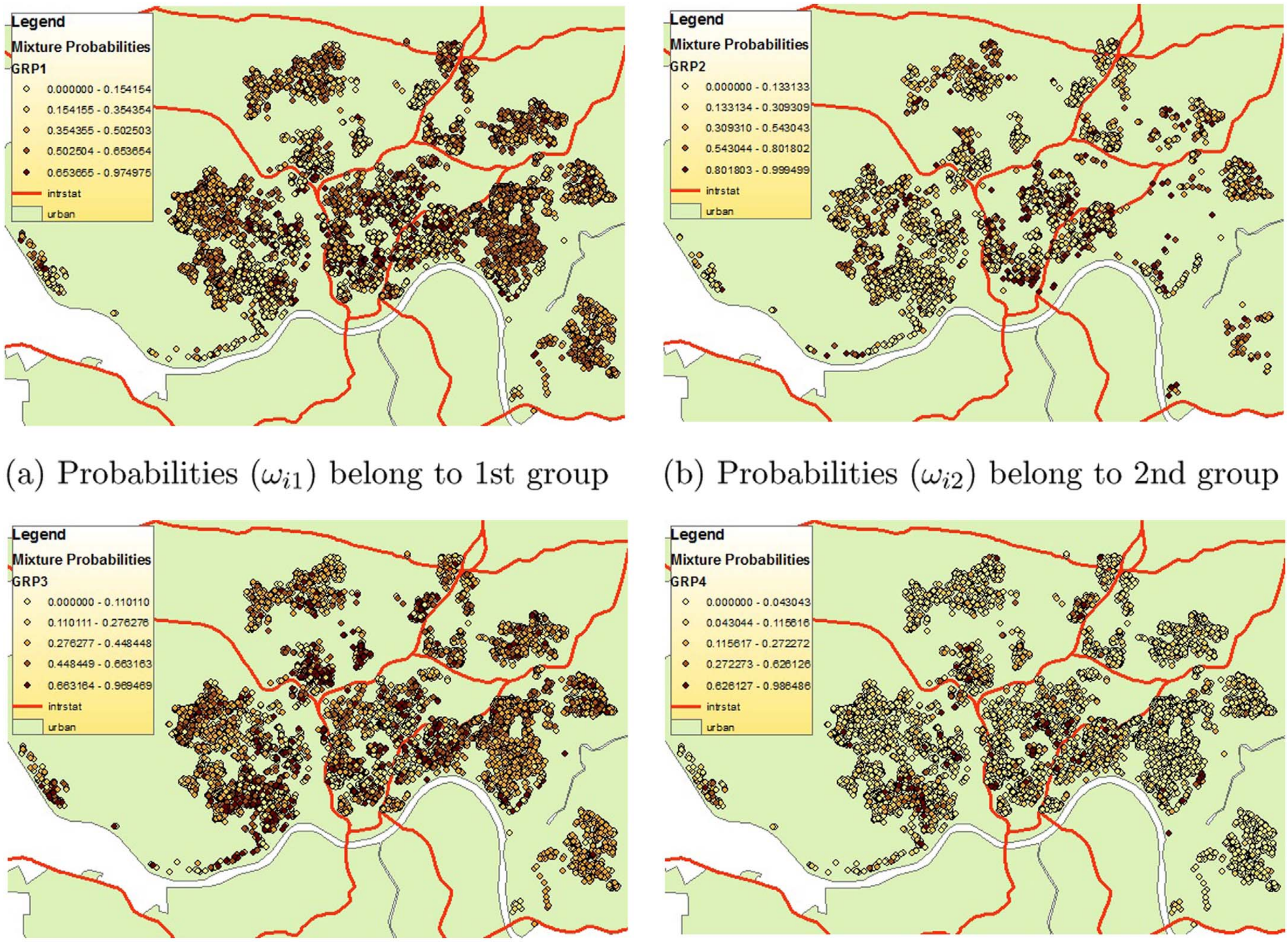
(b) Draws for σ_g^2

(c) Draws for ψ

(d) Draws for π_g

(e) M-H Acceptance Rate

Fig. 4. *Diagnostic Plots - High Signal*: These panels show the post burn-in draws from the conditional distributions of each parameter. The model is stable in this environment with no evidence of switching or non-convergence. These diagnostic plots are from high signal simulations.



(a) Probabilities (ω_{i1}) belong to 1st group (b) Probabilities (ω_{i2}) belong to 2nd group

(c) Probabilities (ω_{i3}) belong to 3rd group (d) Probabilities (ω_{i3}) belong to 4th group

Fig. 5. *Classification probabilities*: The shading in each of these panels is illustrative of the probability that house belongs in the reference group. These probabilities are sourced from the final ω_i vectors from which the group indicator is drawn.

Both the Maximum Likelihood and Bayesian estimation method for the SAR model are not robust to this structure and depend heavily on the parametric family specified by the researcher.

The SAR-M model is a strong, semi-parametric tool that can be used to model data with complex distribution form and nests the standard mixture, SAR, and OLS models as a special case. Future research will focus on extending this structure to more empirically applicable spatial specifications and using the flexibility of mixture distributions to provide new insight regarding the interpretation of direct and indirect effects. Finally, the model presented here is valid only in a cross-sectional environment. Future work will focus on extending it to panel based models with both time static and variant group membership.

Acknowledgements

We would like to thank the Taft Research Center, and the Research Excellence Committee of the Carl H. Lindner College of Business of the University of Cincinnati for providing generous research support. The authors acknowledge comments from James P. LeSage, Don Lacombe and other participants at the Midwest Econometrics Group 2015, and SRSA 2016 conferences.

References

Allenby, G.M., Arora, N., Ginter, J.L., 1998. On the heterogeneity of demand. J. Mark. Res. 35, 384–389.
Anselin, L., Arribas-Bel, D., 2013. Spatial fixed effects and spatial dependence in a single cross-section. Pap. Reg. Sci. 92, 3–17.
Aquaro, M., Bailey, N., Pesaran, M.H., 2015. Quasi maximum likelihood estimation of spatial models with heterogeneous coefficients. USC-INET Research Paper.
Chamberlain, G., 1982. Multivariate regression models for panel data. J. Econ. 18, 5–46.
Chotikapanich, D., Griffiths, W.E., Rao, D.P., 2007. Estimating and combining national income distributions using limited data. J. Bus. Econ. Stat. 25, 97–109.
Cornwall G. 2016. Spatial durbin mixture models. Working Paper.
Dempster, A.P., Laird, N.M., Rubin, D.B., 1977. Maximum likelihood from incomplete data via the em algorithm. J. R. Stat. Soc. Ser. B (Methodol.), 1–38.
Duan, J.a., Guindani, M., Gelfand, A.E., 2007. Generalized spatial dirichlet process models. Biometrika 94, 809–825, (ISSN 00063444).
Epple, D., Quintero, L., Sieg, H., 2014. A New Approach to Estimating Hedonic Pricing Functions for Metropolitan Housing Markets.
Escobar, M.D., West, M., 1995. Bayesian density estimation and inference using mixtures. J. Am. Stat. Assoc. 90, 577–588, (ISSN 0162-1459).
Frühwirth-Schnatter, S., 2004. Estimating marginal likelihoods for mixture and markov switching models using bridge sampling techniques. Econ. J. 7, 143–167.
Frühwirth-Schnatter, S., 2006. Finite Mixture and Markov Switching Models. Springer Science & Business Media.
Gabriel, A.S., Daniels, M.a., Diefendorff, J.M., Greguras, G.J., 2014. Emotional labor actors: a latent profile analysis of emotional labor strategies. J. Appl. Psychol. 100, 863–879, (ISSN 1939-1854).
Gelfand, A.E., Kottas, A., MacEachern, S.N., 2005. Bayesian nonparametric spatial modeling with Dirichlet process mixing. J. Am. Stat. Assoc. 100, 1021–1035, (ISSN 0162-1459).

Geweke, J., 2007. Interpretation and inference in mixture models: simple mcmc works. Comput. Stat. Data Anal. 51, 3529–3550.
Ji, C., Merl, D., Kepler, T.B., West, M., 2009. Spatial mixture modelling for unobserved point processes: examples in immunofluorescence histology. Bayesian Anal. 4, 297–316, (ISSN 19360975).
Kalli, M., Walker, S.G., Damien, P., 2013. Modeling the conditional distribution of daily stock index returns: an alternative bayesian semiparametric model. J. Bus. Econ. Stat. 31, 371–383, (URL http://dx.doi.org/10.1080/07350015.2013.794142).
Kottas, A., Duan, J.a., Gelfand, A.E., 2008. Modeling disease incidence data with spatial and spatio-temporal dirichlet process mixtures. Biom. J. 50, 29–42, (ISSN 03233847).
Kuminoff, N.V., Parmeter, C.F., Pope, J.C., 2010. Which hedonic models can we trust to recover the marginal willingness to pay for environmental amenities? J. Environ. Econ. Manag. 60, 145–160.
LeSage, J., Pace, R.K., 2009. Introduction to Spatial Econometrics. Chapman & Hall/CRC, Boca Raton, FL.
LeSage, J., Vance, C., Chih, Y., 2017. A Bayesian heterogeneous coefficients spatial autoregressive panel data model of retail fuel duopoly pricing. Reg. Sci. Urban Econ. 62, 46-55.
LeSage, J.P., Chih, Y.Y., 2016. Interpreting heterogeneous coefficient spatial autoregressive panel models. Econ. Lett. 142, 1–5.
Pace RK, LeSage JP, Zhu S. 2011. Spatial dependence in regressors and its effect on estimator performance. Available at SSRN 1801241.
Quandt, R.E., 1972. A new approach to estimating switching regressions. J. Am. Stat. Assoc. 67, 306–310.
Quandt, R.E., Ramsey, J.B., 1978. Estimating mixtures of normal distributions and switching regressions. J. Am. Stat. Assoc. 73, 730–738.