On Spatial and Platial Dependence: Examining Shrinkage in Spatially-Dependent Multilevel Models*

Levi John Wolf^{1,2}, Luc Anselin², Daniel Arribas-Bel^{2,3}, and Lee Mobley⁴

¹School of Geographical Sciences University of Bristol

²Center for Spatial Data Science, University of Chicago

³Geographic Data Science Lab, Department of Geography and Planning, University of Liverpool

⁴School of Public Health, Georgia State University

June 16, 2020

Abstract

Multilevel models have been applied to study many geographical processes in epidemiology, economics, political science, sociology, urban analytics, and transportation. They are most often used to express how the effect of a treatment or intervention may vary by geographical group, a form of spatial process heterogeneity. In addition, these models provide a notion of “platial” dependence: observations that are within the same geographical place are modeled as similar to one another. Recent work has shown that spatial dependence can be introduced into multilevel models, and has examined the empirical properties of these models’ estimates. However, systematic attention to the mathematical structure of these models has been lacking. This paper examines a kind of multilevel model that includes both “platial” and “spatial” dependence. Using mathematical analysis, we obtain the relationship between classic multilevel, spatial multilevel, and single-level models. This mathematical structure exposes a tension between a main benefit of multilevel models, estimate shrinkage, and the effects of spatial dependence. We show, both mathematically and empirically, that classic multilevel models may overstate estimate precision and understate estimate shrinkage when spatial dependence is present. This result extends long-standing results in single-level modeling to multilevel models.

*This research was funded in part by NIH Award 7R01CA1266858-06, GeoSpatial Factors and Impacts II, and NSF Award 1657689. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of NIH or the NSF.

Introduction

Geographical processes often exhibit strong differences between places, regions, or locales. In quantitative geography, this kind of *spatial heterogeneity* has long been recognized as important to accurate models (Anselin 1988). One method to address spatial heterogeneity that has seen extensive use in geography is multilevel modeling (Goldstein 2011). Extending earlier work by Goldstein (1986), early discussions of multilevel model in geography (Bondi and Bradford 1990; Jones 1991) led the charge on the “local turn” in geographical analysis (Fotheringham 1997). For example, early arguments that multilevel models allow “relationships to vary from place to place and over time.” (Jones 1991, p. 148) anticipated many other now-common local models in geography (e.g. Brunsdon et al. 1996). Multilevel models remain desirable across a very wide swath of science precisely because of this ability to account for group heterogeneity (Gelman 2015).

As such, multilevel models have been used for a long time in a variety of geographical problem domains. They are used to express or reflect group-wise heterogeneity in the process being modelled, and are useful in reducing the uncertainty of effect estimates while understanding their size (Gelman 2006a). One of the earliest uses of multilevel models came from education (Raudenbush and Bryk 1986; Goldstein 1986), focusing on estimating the “value added” by specific schools, and this work on the geography of education is ongoing (Harris et al. 2007; Browne and Goldstein 2010; Reardon et al. 2019). Studies in epidemiology use multilevel models to estimate “contextual” effects in processes like smoking (Jones and Duncan 1995; Duncan et al. 1996; Moon et al. 2012), mental health (Larsen and Merlo 2005; Chaix, Leyland, et al. 2006), asthma (Wright and S. V. Subramanian 2007), or cancer (Mobley, Kuo, Urato, and S. Subramanian 2010; Mobley, Kuo, Urato, S. Subramanian, et al. 2012). They are used often in the analysis of urban social and economic structure (Dong, Wolf, et al. 2018; Petrovi et al. 2018; Comber et al. 2020) ¹.

Generally, multilevel models in geography are “regional” models because their model parameters change by geographical group (i.e. “region”) (Dong and Harris 2015) in a manner similar to “regimes” regression in spatial econometrics (Anselin 1990; Ertur et al. 2007). This means multilevel models are chiefly powered by the fact that observations within the same group are correlated. When groups are geographical, this represents “patial” dependence (Arcaya et al. 2012), that observations from the same

¹ Consult Diez-Roux (2000) or Pickett and Pearl (2001) for a review of early work in this field

place are correlated. In most study, “patial” dependence is different from “spatial” dependence (Anselin 1988), which usually models the fact that nearby observations depend on one another regardless of regional boundaries (Owen et al. 2015). Both kinds of dependence require different mathematical structures to model. Work introducing spatial dependence structures into multilevel models (Corrado and Fingleton 2012; Arcaya et al. 2012; Dong and Harris 2015; Lacombe, Holloway, et al. 2014; Dong and Wu 2016; Lacombe and McIntyre 2016; Dong, Ma, et al. 2016) has been successful at defining models incorporating both “patial” and “spatial” dependence. This literature has been remarkably successful in providing insight into the practical impact of accounting for both “patial” and “spatial” dependence.

However, this literature has been less successful in defining and discovering *formal* properties of these models. While we have good understandings of the formal properties of classic multilevel models and single-level spatial models, we do not have a good understanding of how their unique behaviors interact. For example, one significant factor driving the use of multilevel models (over similar single-level model specifications) is *shrinkage*: extreme estimates tend to reduce in magnitude generally reduce in their uncertainty (Gelman 2006a). In contrast, simultaneous autoregressive error models (Anselin 1988) are a very common treatment to address dependence in single-level model residuals in spatial econometrics. They are desirable chiefly because they correct for over-confidence in standard linear model estimates while leaving their point estimate unchanged. (Anselin 1988, p. 59)

This means that a spatial multilevel model containing both spatial and multilevel structures will blend two intrinsically-contradictory behaviors. If the multilevel structure dominates, we will see biased but more precise estimates, but if the spatial error model dominates, we will see unbiased but less precise estimates. Without new formal mathematical work describing this trade-off, we can only describe how the tradeoff is made in specific datasets.

Therefore, we focus on the mathematical structure of spatial multilevel models. We build upon results from classic multilevel models to provide the first expression of a *shrinkage matrix* for spatial multilevel models. This matrix provides the structural relationship between single-level spatial models, multilevel models, and their spatial multilevel mixtures. After identifying this mathematical structure and outlining the intrinsic tension discussed above, we explore its empirical ramifications, providing a new diagnostic to understand where, why, and how these two intrinsically-opposed model components trade off.

Spatially-Correlated Multilevel Models

To understand how “spatial” and “patial” dependence work in spatial multilevel models, it helps to first explain multilevel models and spatial effects. Multilevel models sit at the junction of two different kinds of spatial effects in geography. The first involves the idea of *spatial dependence*, that observations in spatial processes tend to be related to other observations nearby. The second effect *spatial heterogeneity*, that processes might be different in different places, tends to dominate the discussions about multilevel models in geography. This is because multilevel models are usually specified within some explicitly-geographical hierarchy collecting individual observations into groups, such as observations at county level grouped into states or observations at neighborhood level grouped into cities. Then, multilevel models allow components of the model, such as the mean, to vary by group, while ensuring that observations within the same group are correlated with one another.

However, early explorations of multilevel models in geography were interested in addressing spatial dependence, too (Jones 1991, p. 148): observations in the same geographical group tend to also be geographically near one another. However, as Arcaya et al. (2012) note, this is not enough: “spatial” dependence means nearby observations are related regardless of their place membership.

To demonstrate this issue, it is sufficient to use a “variance components” or “varying-intercept” multilevel model (Gelman 2005; Browne and Draper 2006). In their simplest form, variance components specifications use a nested hierarchical process to model separate lower- and upper-level variability.² Practically speaking, this empowers spatial analysts to separate the uncertainty related to estimating group-level effects from the uncertainty about the overall response in the model (Gelman 2006a). In geography, groups are typically “regions,” exogenously-determined collections of the observational units known at the start of analysis and fixed throughout. Since regions can be very heterogeneous at many geographical scales, this model specification has long been of interest in geography (Jones 1991).

These models take the form below:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \Delta\boldsymbol{\zeta} + \boldsymbol{\epsilon} \quad (1)$$

²Generally, we refer to “regions” when discussing geographical groups, and these constitute the “upper-level” part of a multilevel model. In contrast, we refer to “observations” when discussing the members of regions, and these constitute the “lower-level” part of a multilevel model.

where $\mathbf{X}$ is an $n \times p$ matrix containing n observations over p covariates about the system response $\mathbf{y}$, with β as the vector of marginal effects relating covariate $\mathbf{x}_k$, $k = 1, 2, \dots, p$ to $\mathbf{y}$. Then, Δ is the “dummy variable” matrix relating the response level, where $\mathbf{X}$ and $\mathbf{y}$ are observed, to the region-level, where a region-level error component, ζ , is modelled.³

To introduce spatial autoregressive effects in the variance components, additional structure must be imposed on ζ and ϵ . Let the covariance of each error component be a function (Ψ, Ω) with two parameters, a spatial autoregressive parameter (λ, θ) and scaling parameter $(\sigma_\epsilon^2, \sigma_\zeta^2)$. With this, a variance components model with generic spatial covariance is:

$$\begin{aligned}\mathbf{y} &= \mathbf{X}\beta + \mathbf{v} \\ \mathbf{v} &= \Delta\zeta + \epsilon \\ \epsilon &\sim \mathcal{N}(0, \Psi(\lambda, \sigma_\epsilon^2)) \\ \zeta &\sim \mathcal{N}(0, \Omega(\theta, \sigma_\zeta^2))\end{aligned}\tag{2}$$

where $\lambda, \sigma_\epsilon^2$ are the response-level spatial autoregressive and scale parameters for component ϵ , and θ, σ_ζ^2 are the spatial autoregressive and scale parameters for the region-level component, ζ . Using this spatial covariance matrix, a distributional model of the response, $\mathbf{y}$, is available:

$$\mathbf{y} \sim \mathcal{N}(\mathbf{X}\beta, \Delta\Omega(\theta, \sigma_\zeta^2)\Delta' + \Psi(\lambda, \sigma_\epsilon^2))\tag{3}$$

Both variance components affect $\mathbf{y}$. This makes the resulting variance-covariance matrix for the model rather complex. But, it can be split into three distinct sources. The first is a within-region (“patial”) dependence derived purely from hierarchical conditioning. This is present due to the Δ terms in the covariance matrix, and is only “spatial” correlation if the groups themselves are spatially-designed. The second factor provides between-region spatial dependence, in that nearby region effects depend on one another. Third, the response-level errors are spatially-dependent, and may impact both within- and between-region covariance. These are represented by Ω and Ψ , respectively.⁴

³This model is sometimes stated with region-level data, $\mathbf{Z}$, and region-level marginal effects γ , but this can be restated into a form equivalent to Eq. 1. We focus only $\mathbf{X}\beta$ to simplify notation and emphasizes the focus of the variance components model: the distinct sources of process variability for groups & observations.

⁴The discussion below requires a few assumptions about these models of spatial dependence. These assumptions are shared by many common models of spatial dependence used in geographical analysis, including both simultaneous and conditional specifications. We detail them in Appendix A.

Mathematical Analysis of the Spatial Multilevel Model

With this specified, we can provide a formal relationship between single-level spatial models, multilevel models, and their spatial-multilevel counterparts. To do this, we follow Jones and Bullen (1994) to provide a closed-form statement of how the estimates from standard single-level models compare to their multilevel counterparts.⁵

Obtaining Shrinkage Matrices for Multilevel Models

In multilevel models, “shrinkage” refers to how the group-level estimates (ζ) change when moving from a single-level to a multilevel model. In the aspatial case, multilevel model estimates “shrink” from their single-level counterparts towards their multilevel estimates. From Jones and Bullen (1994), this can be expressed as an explicit relationship between the multilevel estimate for group j (on left) and the single-level estimate (on right):

$$\mathbf{E}[\hat{\zeta}_{j,mlm}] = \frac{\sigma_z^2}{\sigma_z^2 + \frac{\sigma_e^2}{n_j}} \hat{\zeta}_{j,ols} \quad (4)$$

The fraction in front of the single-level OLS estimate ($\hat{\zeta}_{j,ols}$) is can never be greater than one, since σ_e^2, σ_u^2 and n_j are always positive. So, the expected multilevel estimate is never larger than the corresponding OLS estimate. The fraction preceding the OLS estimate is sometimes called the “reliability” or “estimating precision” for ζ , but this naming is too easy to confuse with standard error. Instead, we adopt the terminology from Rabash et al. (2012) and call this the *shrinkage factor*. Since Eq. 4 is expressed only for group j , we can state an entire *shrinkage matrix* as:

$$\mathbf{E}[\hat{\zeta}_{mlm}] = \left[\left(\mathbf{I}_J \sigma_z^2 + (\Delta' \Delta)^{-1} \sigma_e^2 \right)^{-1} \sigma_z^2 \right] \hat{\zeta}_{ols} \quad (5)$$

In the classic multilevel setup, all shrinkage factors are a function of the upper- and lower-level variance, which are global parameters that apply equally to all observations. The only “place” term in shrinkage is n_j , the number of responses in the focal region j . If all regions have the same number of observations, the shrinkage is uniform across the map. Critically for a spatial setting, the shrinkage

⁵Detail on how this is done is provided in Appendix A. A fully Bayesian treatment through a Gibbs sampler is one straightforward way to obtain the shrinkage matrix, but these models can also be estimated using Expectation Maximization or Restricted Maximum Likelihood methods. They can also be efficiently approximated using Integrated Nested Laplace Approximation or Approximate Variational Inference. However, a mathematical statement of the shrinkage matrix is our point here, not the final estimates themselves.

experienced in estimates for one region is wholly independent of surrounding regions. Neighboring observations or neighboring regions provide no additional information about the regional estimate $\hat{\zeta}_j$, beyond contributing to the estimate of σ_z^2 like any other far-away region. Beyond how groups are constructed, geography plays no further role.

Shrinkage in Spatial Multilevel Models

To acquire a *shrinkage factor* for spatial multilevel models, we need an expression like Eq. 5 that includes spatial dependence in group- and observation-level errors. We will examine one specification (a simultaneous autoregressive structure) and note that the same steps apply for other spatial error covariance specifications.⁶ As discussed by Anselin (1988), estimation of the simultaneous autoregressive spatial error model is a special case of a generalized least squares problem. If the spatial dependence parameter (λ) were known, the regional estimate would be:

$$\hat{\zeta}_{se} = (\Delta' \mathbf{F}' \mathbf{F} \Delta)^{-1} \Delta' \mathbf{F}' \mathbf{F} \mathbf{y} \quad (6)$$

where $\mathbf{F} = \mathbf{I}_N - \lambda \mathbf{W}$. From this estimator and our expression for ζ in the spatial multilevel model, we can define the shrinkage factor in a spatial model.

In both classic and spatial multilevel models, shrinkage is used to describe the model after estimation. So, with an estimate of λ the regional estimate in a spatial multilevel model can be stated:

$$\mathbf{E}[\hat{\zeta}_{smlm}] = \left[\Delta' \mathbf{F}' \mathbf{F} \Delta \sigma_e^{-2} + \Omega(\theta, \sigma_z^2)^{-1} \right]^{-1} \Delta' \mathbf{F}' \mathbf{F} \sigma_e^{-2} \mathbf{y} \quad (7)$$

To define a spatially-dependent multilevel analogue of the shrinkage factor defined by Jones and Bullen (1994), we must rearrange terms. After doing so,⁷ we obtain a shrinkage matrix for spatial multilevel models:

$$\mathbf{E}[\hat{\zeta}_{smlm}] = \left[\left(\mathbf{I}_J \sigma_z^2 + (\Delta' \mathbf{F}' \mathbf{F} \Delta)^{-1} \Omega(\theta)^{-1} \sigma_e^2 \right)^{-1} \sigma_z^2 \right] \hat{\zeta}_{se} \quad (8)$$

⁶Spatial Markov Random Field (SMRF) structures for spatial dependence (Cressie and Wike 2011, §4.2.1) require a separate treatment. Common SMRF specifications reflect different assumptions from simultaneous models of spatial dependence (Wall 2004) and can behave differently from other common kinds of simultaneous autoregressive models (Hodges and Reich 2010). We choose to focus on simultaneous autoregressive models, but any Spatial Markov Random Field with separable covariance (discussed in Appendix A) can use the same derivation.

⁷A full derivation for this is shown in Appendix C.

The matrix in brackets is the $J \times J$ shrinkage matrix for the spatial multilevel variance components model. It blends “patial”/regional dependence with spatial dependence. The leftmost term, $\mathbf{I}_J \sigma_z^2$, is the same as in the aspatial multilevel model. It measures the “substantive” variation in the regional estimates. However, the second term now combines the impact of many distinct spatial processes: the number of responses within a given region (in Δ when $\lambda = 0$), the strength of the “direct” relationship between response-level residuals aggregated by region (from $\Delta \mathbf{F}$), and how the upper-level residuals are spatially dependent (from $\Omega(\theta)$). Unfortunately, the spillover between regions means that there is no simple expression for the shrinkage experienced by a single region like the classic result in Eq. 4. The shrinkage matrix now involves spillovers between groups and between observations in different groups, meaning one group’s spatial multilevel estimate will involve *many* groups’ single-level estimates.

An Empirical Illustration of Spatial Shrinkage

The interplay between spatial dependence and shrinkage is critical. For example, say spatial dependence is included in a model, but is weak in a given empirical example. If this occurs, shrinkage remains the same and Eq. 8 reduces to Eq. 5. However, if spatial dependence is strong, it can affect both how estimates shrink towards zero and how their confidence intervals narrow. If spatial dependence exists at either group or observation level and is *ignored* in a multilevel model, estimates may “narrow” too much but not “shrink” enough. This would mean that ignoring spatial dependence results in more extreme, overconfident estimates than an equivalent model specification with spatial dependence included. This is *substantively different* from the single-level case, where introducing spatial dependence in error terms does not change the resulting estimates.

To show this behaviour in practice, we fit a set of spatial multilevel models in an empirical example. We use data on the complete population of Fee-For-Service Medicare recipients. This data is discussed at length in Mobley, Amaral, et al. (2017)’s study of the implementation of the Medicare Prescription Drug, Improvement, and Modernization Act (MMA). This data presents a natural case study for spatial multilevel models. State-level policy differences and county-level population demographics both are relevant for the prevalence of endoscopy screening. The number of counties in each state also varies widely, so even an aspatial multilevel model will likely show differences from a single-level model. The variables used to model endoscopy screening among the Medicare recipients at the county level are

	Explanation
Outcome:	Percent who ever used endoscopy screening during 2001-2005.
Predictors:	Percent aged 65-74 Percent of sample aged 74-84 Percent of sample aged 85+ Percent of sample Female Percent of sample African American Percent of sample Asian Percent of sample Hispanic Percent of sample American Indian Percent of sample with little or no English language ability Percent of sample Graduate/professional degrees

Table 1: Covariates used in the empirical example, observed at county level over contiguous 48 US states. Originally from Mobley, Amaral, et al. (2017)'s analysis Medicare Fee-for-Service enrollees. Note that no exogenous state-level covariates were included in this analysis for simplicity of comparison of the pooled and un-pooled specifications.

shown in Table 1.

As discussed above, our main purpose is to provide repeatable insights into how these models work. The mathematical structure of shrinkage applies in this (and any other) empirical example or simulation study. Since we examine models that specify geography only in the error covariance in each level, the β effects remain constant. Thus, our discussion will focus on the changes in state-level effects (ζ) while visualizing the extent of spillovers in the spatial multilevel model.⁸

The specifications of these models are summarized in Table 2. To avoid confusing the simultaneous spatial autoregressive error specification with the so-called “spatial lag” model (Anselin 1988), we refer to the SAR-Error specification as an SE specification. Altogether, Table 2 defines twelve possible specifications from combinations of the SE and independent identically-distributed errors at either the response or regional level, alongside multilevel or fixed-effect heterogeneity structures. The four multilevel specifications will tend to shrink their estimates of spatial heterogeneity away from their corresponding response-level error model estimates.

⁸For the multilevel models, four Gibbs samplers are simulated from random starts for 22000 iterations. The spatial parameters were sampled using Metropolis-within-Gibbs sampling. After 2,000 iterations, the chains all converged to the same posterior distribution (Gelman and Rubin 1992), with less than a 1 percent improvement on estimate variance likely from more iterations. Effective size was well over 1000 in all parameters in each chain. The last 10000 iterations were pooled for analysis. Single-level models are fit *via* maximum likelihood. All estimators are distributed in PySAL (Rey and Anselin 2007; Wolf 2018).

Model Description	Abbreviation	Pooling	Distributional model for $\mathbf{y}$
Standard Model	SLM	Full	$\mathbf{y} \sim \mathcal{N}(\mathbf{X}\beta + \alpha, \mathbf{I}_N\sigma_e^2)$
Spatial Fixed Effect	OLS-SFE	None	$\mathbf{y} \sim \mathcal{N}(\mathbf{X}\beta + \Delta\zeta, \mathbf{I}_N\sigma_e^2)$
Spatial Fixed Effect Spatial Error	SE-SFE	None	$\mathbf{y} \sim \mathcal{N}(\mathbf{X}\beta + \Delta\zeta, (\mathbf{F}'\mathbf{F})^{-1}\sigma_e^2)$
Multilevel	MLM	Partial	$\mathbf{y} \sim \mathcal{N}(\mathbf{X}\beta, \Delta\Delta'\sigma_z^2 + \mathbf{I}_N\sigma_e^2)$
Regional SAR-Error	U-SE	Partial	$\mathbf{y} \sim \mathcal{N}(\mathbf{X}\beta, \Delta(\mathbf{G}'\mathbf{G})^{-1}\Delta'\sigma_z^2 + \mathbf{I}_N\sigma_e^2)$
Response SAR-Error	L-SE	Partial	$\mathbf{y} \sim \mathcal{N}(\mathbf{X}\beta, \Delta\Delta'\sigma_z^2 + (\mathbf{F}'\mathbf{F})^{-1}\sigma_e^2)$
Response-SE Region-SE	Dual SE	Partial	$\mathbf{y} \sim \mathcal{N}(\mathbf{X}\beta, \Delta(\mathbf{G}'\mathbf{G})^{-1}\Delta'\sigma_z^2 + (\mathbf{F}'\mathbf{F})^{-1}\sigma_e^2)$

Table 2: Table of example model specifications that include spatial heterogeneity and/or spatial dependence. On the left is a short description, and on the right, the unconditional distributional model for $\mathbf{y}$ is stated like $\mathbf{y} \sim \mathcal{N}(\mu, \Sigma_u + \Sigma_l)$, where μ is the mean, Σ_u is the covariance component pertaining to inter-region “upper-level” dependence, and Σ_l is the “lower-level” component. In this table, $\mathbf{F} = \mathbf{I}_N - \lambda\mathbf{W}$, the lower-level SAR-style covariance factor, and $\mathbf{G} = \mathbf{I}_J - \theta\mathbf{M}$, the upper-level SAR-style covariance factor. The first three models are estimated using Maximum Likelihood, and the remainder use Bayesian methods.

OLS	SFE-SE	MLM	U-SE	L-SE	Dual SE
0.9582	0.7584	0.7435	0.7441	0.7754	0.773
SFE-OLS	U-SMA	L-SMA	Dual SMA	SMASE	SESMA
0.7569	0.744	0.7561	0.7435	0.757	0.7737

Table 3: Mean squared error for the standard linear model with no state-specific effects, the single-level spatial fixed effect model, the single-level spatial fixed-effect model with simultaneous autoregressive spatial errors, and the nine combinations of spatially-dependent multilevel specifications

Results

As expected, the substantive effect estimates (β) are the same between all models. Point estimates change, some estimates change from marginally-significant to statistically insignificant, but no interval estimates for β ever become disjoint between the models. The stability in estimates is expected (Anselin 1988, p. 59), and changes in parameter significance should not be fixated upon unless these changes reflect large swings in estimates in their own right (Gelman and Stern 2006). Finally, the mean squared error of the models (shown in Table 3) is quite consistent, except for the OLS model where heterogeneity is not taken into account. Thus, the models considered are broadly similar on their goodness of fit and β estimates, so long as heterogeneity is taken into account. But, like in many other applications, the degree to which estimates shrink towards zero & become more certain is at issue for multilevel analysis.

Trade-offs between Heterogeneity and Dependence

Fortunately, we can visualize a pattern in the way these specifications work when estimated. Specifications from table 2 tend to fall into one of two groups: those with a SE model of residual correlation at the response level, and those without. In Figure 1, the 95 percent interval estimates for the state-specific terms ($\hat{\zeta}$) across six model specifications are plotted alongside one another. The estimates for models assuming no SE model in the response level are shown in cold-tone colors (aqua, purple, and green). The estimates for models with an SE model at the response level are shown in warm-tone colors (orange, fuchsia, and yellow).

It is helpful to give shorter names to these groups for discussion. Therefore, the models with a response-level SE model are called “thick” models, because they involve a dense, non-diagonal error covariance in the response level. The models with no response-level SE model are called “thin” models, since they either have a block-diagonal response-level covariance matrix or have a diagonal

response-level covariance matrix. For each state, specifications are also arranged left to right, with the three intervals on the left being the “thin” specifications with no response-level SE model and the three intervals on the right being the “thick” specifications with an SE model at response level. Finally, the two single-level specifications (the ‘no pooling’ models) are the spatial fixed-effect (SFE) OLS (aqua, top left in legend & leftmost by state) and the spatial fixed-effect spatial error (SFE-SE) model (orange, bottom left in legend, fourth by state). Thus, altogether, Figure 1 shows six different estimates of ζ (normalized baseline use of endoscopy screening by state) across six model specifications, grouped by color into “thick” and “thin” model sets.

Spatial multilevel estimates retain classic multilevel model behavior

Interpreting Figure 1 requires breaking down points of comparison between models. First, the single-level model specifications (SFE OLS & SFE-SE) have the longest interval estimates throughout. The multilevel specifications have much smaller interval estimates, and tend to be less extreme than their corresponding single-level model estimates. That is, they are more precise, but also biased towards zero. This is shrinkage, and is the expected behavior of all multilevel models.

Shrinkage is very useful when the gain in precision is larger than the shrinkage towards zero. For example, some multilevel estimates may even be statistically significant where the single-level estimates are too noisy. In some states with many counties, like Texas (TX) or California (CA), the interval estimates become dramatically more precise. Further, the two most extreme state-specific estimates, Nevada (NV) on the negative side and Delaware on the positive side (DE), experience much stronger shrinkage when accounting for spatial autoregression. However, some states (such as Indiana, (IN)) have plenty of counties ($n_j = 92$), yet experience much stronger shrinkage under the “thick” specifications than the “thin” specification. We will examine this discrepancy later. For now, it suffices to note that no “thick” estimate is distinct from a “thin” one at the 95% confidence level.

Spatial multilevel models tend to shrink more & tighten less

Beyond this, the “thick” model estimates overlap with zero every time the “thin” estimates do, and overlap in some cases where the “thin” models do not. This means that the “thick” models, those that involve spatial dependence at the response level, are more conservative than the “thin” models that

Figure 1: State-specific estimates for six of the twelve possible models discussed in Table 2. Note the two groups of specifications, those where spatial dependence dominates and those where heterogeneity dominates. The full table showing raw estimates is provided in Appendix D.

	SFE OLS	MLM	U-SE	SFE-SE	L-SE	Dual SE
Moran's I for $\hat{\zeta}$	.332 (.097)	.353 (.095)	.418 (.097)	.357 (.097)	.310 (.098)	.493 (.097)
Upper-Level AR ($\hat{\theta}$)	-	-	.615 (.147)	-	-	.663 (.152)
Lower-Level AR ($\hat{\lambda}$)	-	-	-	.560 (.021)	.596 (.022)	.591 (.021)

Table 4: Spatial dependence in state-level effects across different model specifications. The standard error for the estimator is shown in parentheses.

ignore spatial dependence at the response level. For instance, Texas's "thick" model point estimates are more extreme than those in the "thin" spatial multilevel models. But, the two are still not interval distinct from one another. In nearly all other cases, the "thick" spatial multilevel models are more conservative, yielding estimates with smaller magnitudes and more uncertainty. This suggests that classic multilevel models that ignore response-level spatial dependence may overstate the magnitude of heterogeneity and underestimate their uncertainty.

Examining Spatially-Local Shrinkage

With positive spatial dependence, spatial multilevel models experience similar (and usually stronger) shrinkage towards zero than aspatial multilevel models. Mirroring the classic case, this shrinkage tends to be larger for states with fewer counties. Yet, as noted above with Indiana, spatial multilevel models include more information when shrinking estimates. Indeed, the spatial multilevel models experience *spatially local shrinkage*.

Dependence emerges even when not modeled

The autoregressive parameter estimates (λ , θ) are stable when they are present in the model specification. Every estimate of λ or θ overlaps with all other estimates, regardless of whether the specification is single-level or multilevel. After the model is fit, the estimated spatial autocorrelation in $\hat{\zeta}$ (using Moran's I) is stable and significant. No model has markedly-different spatial autocorrelation in $\hat{\zeta}$. This suggests that spatial dependence in regional effects may emerge regardless of whether the model includes spatial dependence or even considers multilevel structure. It is the impact of spatial dependence on point and interval estimates that matters.

Figure 2: Breakdown of shrinkage matrix values by row, showing diagonal (direct) and off-diagonal (spillover/indirect) influences on $\hat{\zeta}_j$.

Regional spillover shrinkage is more complex

With this, it helps to explicitly examine the shrinkage matrix that arises from Eq. 8. One visualization of this is shown in Figure 2. The black lines in each column represent the aspatial multilevel model's shrinkage factor, which varies between 0 and 1. The remaining colored bars represent the on-diagonal elements of the shrinkage matrix and the off-diagonal elements. On-diagonal elements represent the component of a region's multilevel estimate coming from the region's own single-level estimate. These elements are positive. Off-diagonal elements represent the component of a region's multilevel estimate coming from other regions' single-level estimates. Focusing on Figure 2, the on-diagonal element tends to dominate, but the off-diagonal component can be quite large. Further, negative values can occur in the shrinkage matrix, which is not possible for classic multilevel models.

Therefore, it is important to understand that the shrunk estimate for region j is:

$$\hat{\zeta}_{j,smlm} = \mathbf{S}_{jj}\hat{\zeta}_{j,se} + \sum_{k \neq j}^J \mathbf{S}_{jk}\hat{\zeta}_{k,se} \quad (9)$$

The contribution of region k to the spatial multilevel estimate for region j is dependent on the sign of $\hat{\zeta}_{k,se}$ and the shrinkage matrix at $\mathbf{S}_{jk}$. Assuming all $\hat{\zeta}_{se}$ were positive, then a negative shrinkage factor ($\mathbf{S}_{kj}$) would imply that k brings j 's estimate towards zero. However, not all $\hat{\zeta}_{se}$ are positive in practice. If $\hat{\zeta}_{j,se}$ is positive and $\mathbf{S}_{jk}\hat{\zeta}_{k,se}$ is positive, then region k actually pushes the spatial multilevel estimate for j away from zero. Thus, other regions can cause either *shrinkage* or *growth* for region j , whereas no shrinkage factor can give “growth” in a classical multilevel model.

Figure 3: Maps & shrinkage matrices for spatial multilevel models. The top row shows the shrinkage matrix. Subsequent rows focus on one state each. Note that the classic multilevel model shrinks only with respect to the focal state (marked with a star), yet all spatial multilevel models experience regional spillovers affecting the spatial multilevel estimate.

Shrinkage is driven by factors local to each region

The structure of the shrinkage matrix becomes more clear when mapped. We map the spatial distribution of shrinkage spillovers in Figure 3 for five states and compare it to the aspatial shrinkage in the classical multilevel model. We also show the shrinkage matrices directly in the top row of Figure 3. Note that on the leftmost column, no regional spillovers are possible in the classic multilevel model. However shrinkage is driven both by the focal state and by surrounding states in the spatial multilevel model. Shrinkage spillovers from second-order (and higher) neighbors are present, but the size of contribution decreases quickly with distance. The regions immediately surrounding each observation contribute the most to its shrinkage and, in some cases (e.g. Delaware (DE)), approach the impact of the focal state alone.

Finally, it is critical to note that the shrinkage matrix for spatial multilevel models is not symmetric. This is an interesting property. First, regions have different sizes, so big regions exert more influence. Second, regions have different connectivity, so well-connected regions stand to be influenced by more nearby regions. Third, the “spillover” due to response-level dependence will change at the border of each region, with some borders containing many nearby responses and others containing few (if any). Thus, this is another result of how the three types of dependence, (1) “patial” and spatial dependence in the (2) regional or (3) response geography, combine to change how shrinkage works in spatial multilevel models.

Discussion & Conclusion

Multilevel models are a powerful tool for analyzing geographical data. Recent work in the multilevel modeling literature acknowledges the limitations of multilevel models to model spatial (rather than “patial”) dependence, and has aimed to introduce spatial dependence into multilevel structures. This blends two kinds of models with very different behaviors. Multilevel models are generally used because they narrow interval estimates while becoming slightly biased. In contrast, spatial models are generally used because they widen interval estimates to account for spatial dependence, but yield unbiased estimates. These two contradictory behaviors combine in complex ways for spatial multilevel models.

We show that shrinkage and interval estimate widths are substantially different between aspatial and

spatial multilevel models. Models with strong spatial dependence tend to yield *more biased, less certain* estimates. Their estimates shrink more strongly towards zero and do not see the same improvements in certainty. Viewed in the other direction, *ignoring* spatial dependence in multilevel models can result in more extreme estimates that have too narrow confidence intervals. Thus, the spatial multilevel model can have substantially different behaviour than either a spatial model or a multilevel model would alone.

At a higher level, our goal in providing a thorough mathematical analysis of model structures is to show that further work is needed to build a practical understanding about how spatially-dependent multilevel models work. Through mathematical structure latent in the models geographers use, we can build transferable insight to more complex geographical models. Thus, we stress the value of mathematical analysis of model specifications for future work.

There are a few clear targets for future mathematical analysis. The shrinkage matrix is useful to characterize, map, and understand how regions affect one another in spatial multilevel models. But, it only characterizes the regional spillovers. In addition to our work, a deeper mathematical understanding of multi-scale leverage in spatially-correlated multilevel models would help tremendously. Substantively, it would allow us to identify which estimates are influenced by “local” spatial structure versus “regional” structure in these explicitly multi-scale models. Work extending earlier results in econometrics on leverage, spillovers, & observation influence (Haining 1994; Lesage and Fischer 2008) to current work on multilevel diagnostics (Loy and Hofmann 2014) will certainly prove fruitful here.

We also hope to show that intuition from simpler models does not necessarily transfer directly to more complex mixtures of models. Spatial error dependence usually does not induce significant changes in single-level estimates, and does not do so asymptotically. Indeed, the spatial lag model tends to garner more attention due to its strong implications for estimates and model interpretation. In spite of this, spatial dependence in the upper-level residuals still reflects “substantive” dependence, changing shrinkage and certainty. Of course, this is a matter of perspective in model elucidation, which makes the choice of specification much more complicated.

Beyond the simple “error” dependence considered here, spatial multilevel models with endogenous spatial lag specifications will have even more complex shrinkage behavior. The analysis shown here must be extended to the pathbreaking studies of Arcaya et al. (2012), Corrado and Fingleton (2012), Lacombe, Holloway, et al. (2014), Dong and Harris (2015), and Lacombe and McIntyre (2016). Results that can apply to many different specifications of spatial multilevel model may need to focus on mathe-

mathematical structures that are shared across specifications, rather than structures that are specific to the model at hand. This is difficult, but necessary.

The integration of the historically-distinct literatures on models of spatial heterogeneity with models of spatial dependence clearly stands to benefit spatial analysis. Multilevel modeling and models with spatial dependence have venerable histories their own domains. They both yield useful results in a large variety of applications. Their mathematically-rigorous combination will surely ground even more useful and realistic analytical methods for spatial systems. Through better mathematical insight on the meaning and behavior of model structures, the impacts of these structures can be better and more precisely understood. We hope that future work continues in this vein.

References

- Anselin, L. (1988). *Spatial Econometrics: Methods and Models*. Dordrecht: Kluwer.
- Anselin, L. (1990). Spatial Dependence and Spatial Structural Instability in Applied Regression Analysis. *Journal of Regional Science*, 30(2), 185–207.
- Arcaya, M., Brewster, M., Zigler, C. M., & Subramanian, S. (2012). Area variations in health: A spatial multilevel modeling approach. *Health & Place. Infectious Insecurities*, 18(4), 824–831. doi:10.1016/j.healthplace.2012.03.010
- Bondi, L. & Bradford, M. (1990). Applications of Multi-Level Modelling to Geography. *Area*, 22(3), 256–263.
- Browne, W. & Draper, D. (2006). A comparison of Bayesian and likelihood-based methods for fitting multilevel models. *Bayesian Analysis*, 1(3), 473–514.
- Browne, W. & Goldstein, H. (2010). MCMC Sampling for a Multilevel Model With Nonindependent Residuals Within and Between Cluster Units. *Journal of Educational and Behavioral Statistics*, 35(4), 453–473. doi:10.3102/1076998609359788
- Brunsdon, C., Fotheringham, A. S., & Charlton, M. (1996). Geographically Weighted Regression: A method for exploring spatial nonstationarity. *Geographical Analysis*, 28(4).
- Carpenter, B., Gelman, A., Hoffman, M., Lee, D., Goodrich, B., Betancourt, M., . . . Riddell, A. (2016). Stan: A probabilistic programming language. *Journal of Statistical Software*, *In press*.

- Chaix, B., Leyland, A. H., Sabel, C. E., Chauvin, P., Råstam, L., Kristersson, H., & Merlo, J. (2006). Spatial clustering of mental disorders and associated characteristics of the neighbourhood context in Malmö, Sweden, in 2001. *Journal of Epidemiology and Community Health*, 60(5), 427–435.
- Chaix, B., Merlo, J., Subramanian, V., & Chauvin, P. (2005). Comparison of a spatial perspective with the multilevel analytical approach in neighborhood studies: The case of mental and behavioral disorders due to psychoactive substance use in Malmö, Sweden, 2001. *American Journal of Epidemiology*, 162(2), 171–182.
- Comber, S., Arribas-Bel, D., Singleton, A., Dong, G., & Dolega, L. (2020). Building Hierarchies of Retail Centers Using Bayesian Multilevel Models. *Annals of the American Association of Geographers*, 110(4), 1150–1173. doi:10.1080/24694452.2019.1667219
- Corrado, L. & Fingleton, B. (2012). WHERE IS THE ECONOMICS IN SPATIAL ECONOMETRICS?*. *Journal of Regional Science*, 52(2), 210–239. doi:10.1111/j.1467-9787.2011.00726.x
- Cressie, N. & Wikle, C. K. (2011). *Statistics for spatio-temporal data*. New York: Wiley.
- Diez-Roux, A. V. (2000). Multilevel Analysis in Public Health Research. *Annual Review of Public Health*, 21(1), 171–192. doi:10.1146/annurev.publhealth.21.1.171
- Dong, G. & Harris, R. (2015). Spatial autoregressive models for geographically hierarchical data structures. *Geographical Analysis*, 47(2), 173–191.
- Dong, G., Ma, J., Harris, R., & Pryce, G. (2016). Spatial Random Slope Multilevel Modeling Using Multivariate Conditional Autoregressive Models: A Case Study of Subjective Travel Satisfaction in Beijing. *Annals of the American Association of Geographers*, 106(1), 19–35. doi:10.1080/00045608.2015.1094388
- Dong, G., Wolf, L. J., Alexiou, A., & Arribas-Bel, D. (2018). Inferring neighbourhood quality with property transaction records by using a locally adaptive spatial multi-level model. *Computers, Environment and Urban Systems*. doi:10.1016/j.compenvurbsys.2018.09.003
- Dong, G. & Wu, W. (2016). Schools, land markets and spatial effects. *Land Use Policy*, 59, 366–374.
- Duncan, C., Jones, K., & Moon, G. (1996). Health-related behaviour in context: A multilevel modelling approach. *Soc. Sci. Med.* 42(6), 817–830.
- Ertur, C., Gallo, J. L., & LeSage, J. P. (2007). The European regional convergence process, 1980-1995: Do spatial dependence and spatial heterogeneity matter? *Rev. Reg. Stud.* 37, 82–108.

- Fotheringham, A. S. (1997). Trends in quantitative methods I: Stressing the local. *Progress in Human Geography*, 21(1), 88–96. doi:10.1191/030913297676693207
- Gelman, A. (2005). Analysis of variance?why it is more important than ever. *The Annals of Statistics*, 33(1), 1–53. doi:10.1214/009053604000001048
- Gelman, A. (2006a). Multilevel (Hierarchical) Modeling: What It Can and Cannot Do. *Technometrics*, 48(3), 432–435.
- Gelman, A. (2006b). Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper). *Bayesian Analysis*, 1(3), 515–534.
- Gelman, A. (2015). The Connection Between Varying Treatment Effects and the Crisis of Unreplicable Research: A Bayesian Perspective. *Journal of Management*, 41(2), 632–643. doi:10.1177 / 0149206314525208
- Gelman, A. & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, 7(4), 457–472.
- Gelman, A. & Stern, H. (2006). The difference between “significant” and “not significant” is not itself statistically significant. *The American Statistician*, 60(4), 328–331.
- Goldstein, H. (1986). Multilevel mixed linear model analysis using iterative generalized least squares. *Biometrika*, 73(1), 43–56.
- Goldstein, H. (2011). *Multilevel Statistical Models*. John Wiley & Sons.
- Haining, R. (1994). Diagnostics for Regression Modeling in Spatial Econometrics. *Journal of Regional Science*, 34(3), 325–341.
- Harris, R., Johnston, R., & Burgess, S. (2007). Neighborhoods, Ethnicity and School Choice: Developing a Statistical Framework for Geodemographic Analysis. *Population Research and Policy Review*, 26(5), 553–579. doi:10.1007/s11113-007-9042-9
- Hodges, J. S. & Reich, B. J. (2010). Adding spatially-correlated errors can mess up the fixed effect you love. *The American Statistician*, 64(4), 325–334.
- Hoffman, M. D. & Gelman, A. (2014). The No-U-Turn Sampler: Adaptively Setting Path Lengths in Hamiltonian Monte Carlo. *Journal of Machine Learning Research*, 15(1), 1593–1623.
- Huang, A. & Wand, M. P. (2013). Simple marginally noninformative prior distributions for covariance matrices. *Bayesian Analysis*, 8(2), 439–452.

- Jones, K. (1991). Specifying and estimating multilevel models for geographic research. *Transactions of the Institute of British Geographers*, 16, 148–159.
- Jones, K. & Bullen, N. (1994). Contextual Models of Urban House Prices: A Comparison of Fixed- and Random-Coefficient Models Developed by Expansion. *Economic Geography*, 70(3), 252–272. doi:10.2307/143993
- Jones, K. & Duncan, C. (1995). Individuals and their ecologies: Analysing the geography of chronic illness within a multilevel modelling framework. *Health & Place*, 1(1), 27–40. doi:10.1016/1353-8292(95)00004-6
- Kass, R. E. & Natarajan, R. (2006). A default conjugate prior for variance components in generalized linear mixed models (comment on article by Browne and Draper). *Bayesian Analysis*, 1(3), 535–542.
- Kelejian, H. H. & Prucha, I. R. (1998). A generalized spatial two-stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbances. *J. Real Estate Fin. Econ.* 17(1), 99–121.
- Lacombe, D. J., Holloway, G. J., & Shaughnessy, T. M. (2014). Bayesian estimation of the spatial Durbin error model with an application to voter turnout in the 2004 presidential election. *International Regional Science Review*, 37(3), 298–327.
- Lacombe, D. J. & McIntyre, S. G. (2016). Local and global spatial effects in hierarchical models. *Applied Economics Letters*, 23(16), 1168–1172. doi:10.1080/13504851.2016.1142645
- Larsen, K. & Merlo, J. (2005). Appropriate Assessment of Neighborhood Effects on Individual Health: Integrating Random and Fixed Effects in Multilevel Logistic Regression. *American Journal of Epidemiology*, 161(1), 81–88. doi:10.1093/aje/kwi017
- LeSage, J. P. (1997). Bayesian estimation of spatial autoregressive models. *International Regional Science Review*, 20(1-2), 113–129.
- Lesage, J. P. & Fischer, M. M. (2008). Spatial Growth Regressions: Model Specification, Estimation and Interpretation. *Spatial Economic Analysis*, 3(3), 275–304. doi:10.1080/17421770802353758
- LeSage, J. P. & Pace, R. K. (2009). *Introduction to Spatial Econometrics*. Boca Raton, FL: CRC Press.
- Lindgren, F. & Rue, H. (2015). Bayesian spatial modelling with R-INLA. *Journal of Statistical Software*, 63(19), 1–25.

- Lindley, D. V. & Smith, A. F. M. (1972). Bayes estimates for the linear model. *Journal of the Royal Statistical Society. Series B (Methodological)*, 34(1), 1–41.
- Loy, A. & Hofmann, H. (2014). **HLMdiag** : A Suite of Diagnostics for Hierarchical Linear Models in R. *Journal of Statistical Software*, 56(5). doi:10.18637/jss.v056.i05
- Mobley, L., Amaral, P., Kuo, T.-M., Zhou, M., & Bose, S. (2017). Medicare modernization and diffusion of endoscopy in FFS medicare. *Health Economics Review*, 7(1), 13. doi:10.1186/s13561-017-0147-5
- Mobley, L., Kuo, T.-M., Urato, M., & Subramanian, S. (2010). Community contextual predictors of endoscopic colorectal cancer screening in the USA: Spatial multilevel regression analysis. *International Journal of Health Geographics*, 9(1), 1.
- Mobley, L., Kuo, T.-M., Urato, M., Subramanian, S., Watson, L., & Anselin, L. (2012). Spatial Heterogeneity in Cancer Control Planning and Cancer Screening Behavior. *Annals of the Association of American Geographers*, 102(5), 1113–1124. doi:10.1080/00045608.2012.657494
- Moon, G., Pearce, J., & Barnett, R. (2012). Smoking, Ethnic Residential Segregation, and Ethnic Diversity: A Spatio-temporal Analysis. *Annals of the Association of American Geographers*, 102(5), 912–921. doi:10.1080/00045608.2012.657497
- Natarajan, R. & Kass, R. E. (2000). Reference Bayesian methods for Generalized Linear Mixed Models. *Journal of the American Statistical Association*, 95(449), 227–237.
- Neal, R. (2011). MCMC Using Hamiltonian Dynamics. In S. Brooks, A. Gelman, G. Jones, & X.-L. Meng (Eds.), *Handbook of Markov Chain Monte Carlo*. CRC Press.
- Owen, G., Harris, R., & Jones, K. (2015). Under examination: Multilevel models, geography and health research. *Progress in Human Geography*, 40(3), 394–412. doi:10.1177/0309132515580814
- Petrovi, A., van Ham, M., & Manley, D. (2018). Multiscale Measures of Population: Within- and between-City Variation in Exposure to the Sociospatial Context. *Annals of the American Association of Geographers*, 108(4), 1057–1074. doi:10.1080/24694452.2017.1411245
- Pickett, K. E. & Pearl, M. (2001). Multilevel Analyses of Neighborhood Socioeconomic Context and Health Outcomes: A Critical Review. *Journal of Epidemiology and Community Health*, 55(2), 111–122.
- Polson, N. G. & Scott, J. G. (2012). On the half-Cauchy prior for a global scale parameter. *Bayesian Analysis*, 7(4), 887–902.

- Rabash, J., Charlton, C., Browne, W., Healy, M., & Cameron, B. (2012). *A User's Guide to MLwiN (v2.30)*. London: University of Bristol.
- Raudenbush, S. & Bryk, A. S. (1986). A hierarchical model for studying school effects. *Sociology of Education*, 59(1), 1–17.
- Reardon, S. F., Kalogrides, D., & Shores, K. (2019). The Geography of Racial/Ethnic Test Score Gaps. *American Journal of Sociology*, 124(4), 1164–1221. doi:10.1086/700678
- Rey, S. J. & Anselin, L. (2007). PySAL: A Python library of spatial analytical methods. *The Review of Regional Studies*, 37(1), 5–27.
- Ripley, B. D. (1981). *Spatial statistics*. New York, New York, USA: Wiley.
- Robert, C. (2007). *The Bayesian choice: From decision-theoretic foundations to computational implementation*. Springer Science & Business Media.
- Savitz, N. V. & Raudenbush, S. (2009). Exploiting spatial dependence to improve measurement of neighborhood social processes. *Sociological Methodology*, 39(1), 151–183.
- 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(2), 311–324.
- Wolf, L. J. (2018). Pysal/spvcm: 0.2.0: Adding spatial locally-varying coefficient models. Zenodo. doi:10.5281/zenodo.1450945
- Wolf, L. J., Anselin, L., & Arribas-Bel, D. (2018). Stochastic Efficiency of Bayesian MCMC in Spatial Econometric Models: An Empirical Comparison. *Geographical Analysis*, 50(1), 97–119.
- Wright, R. J. & Subramanian, S. V. (2007). Advancing a multilevel framework for epidemiologic research on asthma disparities. *Chest*, 132(5 Suppl), 757S–769S. doi:10.1378/chest.07-1904

A Appendix: Priors & Conditional Posteriors for the Spatially-Correlated Multilevel Model

The models discussed in the main text are estimated using a Metropolis-within-Gibbs approach, owing to (LeSage 1997). The Gibbs steps are simple to obtain, since the model structure combines with the judiciously-chosen priors to provide conditionally-conjugate posterior distributions for all but two of the parameters. The spatial autoregressive parameters, λ and θ have no conditionally-conjugate choice for their prior, and so do not yield a recognized distribution in their conditional posterior. Thus, they must be sampled in another way, such as Metropolis-Hastings. We discuss the configuration and operation of the sampler at length in the main text.

It is important to note that our choice of a Bayesian analysis does not imply this is the only method possible. For instance, Savitz and Raudenbush (2009) examines Expectation Maximization for related specifications. For the uncorrelated case, there are high-quality existing comparisons between Restricted Maximum Likelihood and Bayesian estimation methods (Browne and Draper 2006), as well as comparisons of exact & approximate estimation methods (Lindgren and Rue 2015) in spatial models. The development of multiple estimators is not the intent of this manuscript, however. Instead, this paper is focused on the novel mathematical and empirical behavior of shrinkage in a class of spatial multilevel models, so we use Bayesian analysis as only one of many possible methods to estimate these models.

Separable Simultaneous Spatial Correlation and Scale

In addition to the way conditional likelihoods can be defined in Section , the model structure can be further exploited. Priors for the scale parameters and the spatial correlation parameters are independent in any level of the hierarchy. However, there is no guarantee that the conditional posterior for these terms will be conjugate or separable depending on the structure of Ω or Ψ . Referring to generic spatial correlation and scale parameters, (δ, σ_*^2) , at any level of a variance components model, it is necessary that $\Psi(\delta, \sigma_*^2)$ be “separable” to ensure that the posterior of σ_*^2 remains conjugate. Practically, separability means the spatial correlation function $\Psi(\delta, \sigma_*^2)$ must satisfy:

$$\Psi(\delta, \sigma_*^2) = f(\delta)\sigma_*^2 \tag{10}$$

where f is an arbitrary single-parameter function. If this does not hold, the conditional posteriors cannot be conjugate, even though the priors are independent univariate distributions. Fortunately, separability holds for many covariance specifications, including spatial moving average (SMA), simultaneous autoregressive error (SAR) structures:

$$\begin{aligned}\Psi(\delta, \sigma_*^2)_{SMA} &= (\mathbf{I} + \delta \mathbf{W})(\mathbf{I} + \delta \mathbf{W})' \mathbf{I} \sigma_*^2 \\ \Psi(\delta, \sigma_*^2)_{SAR} &= ((\mathbf{I} - \delta \mathbf{W})'(\mathbf{I} - \delta \mathbf{W}))^{-1} \mathbf{I} \sigma_*^2\end{aligned}\tag{11}$$

where $\mathbf{I}$ is an identity matrix of the appropriate rank and $\mathbf{W}$ is a row-standardized spatial weighting matrix reflecting the spatial relationships between observations at the appropriate level. Note that the covariance matrix for the SMA error specification will be much sparser than that for the SAR error specification; the matrix in the SMA specification is at most as sparse as $\mathbf{W}$, but the matrix inverse in the SAR specification results in a dense covariance matrix. Indeed, the SMA specification only allows for spatial dependence between immediately adjacent neighbors, while the SAR specification models dependence as decaying with proximity (Kelejian and Prucha 1998).

Separability ensures that the covariance components have distinct impacts on the joint posterior covariance matrix.⁹ As stated above, a wide class of covariance specifications are decomposable in this way, such as the disc error specification explored in Ripley (1981), the distance-based covariance specification in Chaix, Merlo, et al. (2005), the z -transformed correlation matrix of Browne and Goldstein (2010), or the SMA-like specification of Corrado and Fingleton (2012). Separability is desirable, since specifying a full-rank covariance matrix prior on a joint $\Psi(\delta, \sigma_*^2)$ in each level would introduce significant model complexity and may result in unintentionally-informative priors (Natarajan and Kass 2000; Kass and Natarajan 2006; Huang and Wand 2013). In concert with the hierarchical conditioning, a separable covariance with univariate scale and autoregressive parameters provides unique conditional posteriors in each level of the spatially-correlated variance components model.

Obtaining a Gibbs Sampler

To define a Gibbs sampler that can estimate spatial-correlated variance components models, it is useful to define the terms used in deriving Gibbs samplers. For a model with parameters $\mu_j \in \mu$, the likelihood

⁹If the covariance model requires multivariate spatial or aspatial components, then a multivariate covariance function $\Psi(B, \Sigma_*)$ must be reducible to $\Psi(B)\Sigma_*$ for separability to hold.

of $\mathbf{y}$ is defined conditional on parameters μ , $\mathcal{L}(\mathbf{y}|\mu)$. In addition, parameters in μ may depend on more intermediate parameters (i.e. hierarchical parameters) with conditional priors $P(\mu_j|\neg\mu_j)$, while others will have independent, unconditional priors, $P(\mu_l)$.¹⁰ Together, the full conditional posterior distribution of parameter μ_j be found using Bayes's Theorem:

$$P(\mu_j|\mathbf{y}, \neg\mu_j) \propto \mathcal{L}(\mathbf{y}|\mu) \times P(\mu_{-j}|\mu_j) \times P(\mu_j) \quad (12)$$

In addition, distributions and terms encountered during conjugation that are constant with respect to μ_j may be dropped, since Eq. 12 is only needed up to a normalizing constant.

We opt for a Metropolis-within-Gibbs sampler, a workhorse of Bayesian spatial econometrics (LeSage 1997). This choice constrains the possible form of prior distributions, since Gibbs sampling requires *conditionally-conjugate* priors to work efficiently. As more methods for sampling Bayesian models become available (Neal 2011; Carpenter et al. 2016), these specifications might be adapted to use more robust choices of prior, such as the half-Cauchy priors considered by Gelman (2006b) and Polson and Scott (2012). As it stands, though, many cutting-edge samplers that allow for this generality are not yet computationally-efficient in generating the same amount of information about parameters, for even modest problem sizes (Wolf et al. 2018). Thus, we retain conventional conjugate priors, which often have acceptable properties in practice.

Despite appearances, the complexity of a multilevel model is manageable. In the multilevel hierarchy, lower levels of the hierarchy can be simplified by conditioning on the next-highest level of the hierarchy (e.g. Robert 2007, §10.2). This ensures that the model for each level is tractable. Instead of the unconditional response-level distributional model stated in Eq. 3, we will work with model of $\mathbf{y}$ after conditioning on ζ . This conditioning simplifies covariance structures in each level; the response-level likelihood only involves the response-level spatially-correlated covariance matrix, $\Psi(\lambda, \sigma_e^2)$, after conditioning on ζ . Before showing this, it is important to distinguish between the hierarchical prior for ζ , denoted $P(\zeta|\theta, \sigma_z^2)$, and the full conditional for ζ , stated as $P(\zeta|\mathbf{y}, \sigma_z^2, \theta)$. The former is inherent in the model specification; the latter is required in order to construct estimators and depends entirely on the specification of model priors. Both will serve distinct purposes in a Gibbs sampler used for estimation.

With this distinction recognized, the hierarchical prior for ζ is a multivariate normal distribution with

¹⁰The term $\neg\mu_j$ is used as shorthand to indicate all parameters except for μ_j .

a zero mean and upper-level covariance, $\Omega(\theta, \sigma_z^2)$. This is our “region-level” model, contrasting with our “response-level” model for y . Stating both the response- and region-level distributional models:

$$P(y|\beta, \zeta, \lambda, \sigma_e^2) \sim \mathcal{N}(\mathbf{X}\beta + \Delta\zeta, \Psi(\lambda, \sigma_e^2)) \quad (13)$$

$$P(\zeta|\theta, \sigma_z^2) \sim \mathcal{N}(0, \Omega(\theta, \sigma_z^2)) \quad (14)$$

Stating $P(y|\zeta, \dots)$ in the response level implies conditioning on all of the rest of the parameters, e.g. θ and σ_z^2 , as well. Since we now “know” ζ when stating y , the response covariance is simplified, and ζ enters the mean predictor for y .

With this in mind, we will analyze conventional conditionally-conjugate priors. The upper-level model of region outcomes, $P(\zeta|\theta, \sigma_z^2)$, serves as a prior model of the unobserved group effects, so ζ has no additional prior. The other priors are chosen from common conditionally-conjugate families with conventional parameterizations:

$$\begin{aligned} \beta &\sim \mathcal{N}(\mu_\beta, \Sigma_{\beta 0}) \\ \sigma_e^2 &\sim IG(a_e, b_e) \\ \sigma_z^2 &\sim IG(a_z, b_z) \end{aligned} \quad (15)$$

The priors for the spatial parameters (λ, θ) are left unspecified, but they are assumed to satisfy two conditions. First, they are functions of only themselves, their hyperparameters, and may include properties of the spatial weighting matrix over which they apply. This allows for the priors to be any well-behaved independent prior that mixes into a distinct conditional posterior for the parameter. In addition, the spatial parameter priors are proper probability distributions, so they integrate to 1.

The Conditional Posterior for Region-Level Effects

Lindley and Smith (1972) provides a lucid discussion of the mathematical strategies used to obtain estimators in multilevel models. At a high level, the analysis of marginal effects like β or ζ in Eq. 1 involves mixing together the prior and relevant components of the likelihood, recognizing a pattern in the so-called “kernel” of the distribution, and then completing the square within this Gaussian kernel. For $P(\zeta|y, \dots)$, the full conditional posterior for ζ involves the mixture of the response-level likelihood

and region-level hierarchical prior. It is this update of the Gibbs sampler that communicates between the two levels of the model:

$$P(\zeta|\mathbf{y}, \dots) \propto \mathcal{L}(\mathbf{y}|\dots) \times P(\zeta|\theta, \sigma_z^2) \quad (16)$$

First, the kernel is expanded and terms are grouped into components that contain our target parameter ζ and those that do not:

$$\propto \exp \left[-\frac{1}{2} (\mathbf{y} - \mathbf{X}\beta - \Delta\zeta)' \Psi(\lambda, \sigma_e^2)^{-1} (\mathbf{y} - \mathbf{X}\beta - \Delta\zeta) \right] \times \exp \left[-\frac{1}{2} \zeta' \Omega(\theta, \sigma_z^2)^{-1} \zeta \right] \quad (17)$$

$$\propto \exp \left[-\frac{1}{2} \left(\zeta \left(\Delta' \Psi(\lambda, \sigma_e^2)^{-1} \Delta + \Omega(\theta, \sigma_z^2)^{-1} \right) \zeta - \zeta' \Delta' \Psi(\lambda, \sigma_e^2)^{-1} (\mathbf{y} - \mathbf{X}\beta) \right. \right. \\ \left. \left. - (\mathbf{y} - \mathbf{X}\beta)' \Psi(\lambda, \sigma_e^2)^{-1} \Delta \zeta + (\mathbf{y} - \mathbf{X}\beta)' \Psi(\lambda, \sigma_e^2)^{-1} (\mathbf{y} - \mathbf{X}\beta) \right) \right] \quad (18)$$

Fortunately, terms that are constant with respect to ζ (like the last term in Eq. 18) can be dropped, since the posterior distribution for ζ in Eq. 16 only needs to be known up to a constant. Further, we can make the pattern in Eq. 18 clearer using a substitution:

$$z = \Delta' \Psi(\lambda, \sigma_e^2)^{-1} (\mathbf{y} - \mathbf{X}\beta) \quad \Sigma_\zeta = \left(\Delta' \Psi(\lambda, \sigma_e^2)^{-1} \Delta + \Omega(\theta, \sigma_z^2)^{-1} \right)^{-1} \quad (19)$$

With this substitution and dropping constant terms, we obtain:

$$\propto \exp \left[\zeta' \Sigma_\zeta^{-1} \zeta - \zeta' z - z' \zeta \right] \quad (20)$$

Then, to complete the square, we can introduce $\pm z' \Sigma_\zeta z$. This term is also constant with respect to ζ , but we can use its positive component to complete the square and obtain the kernel of a normal distribution for our conditional posterior:

$$\propto \exp \left[\left(\zeta' \Sigma_\zeta^{-1} \zeta - \zeta' z - z' \zeta + z' \Sigma_\zeta z \right) - z' \Sigma_\zeta z \right] \quad (21)$$

$$\propto \exp \left[(\zeta - \Sigma_\zeta z)' \Sigma_\zeta^{-1} (\zeta - \Sigma_\zeta z) \right] \quad (22)$$

$$P(\zeta|\mathbf{y}, \dots) \propto \mathcal{N}(\Sigma_\zeta z, \Sigma_\zeta) \quad (23)$$

Using a similar set of techniques, sampling statements for all other conditional posteriors can be

obtained. However, for spatial autoregressive parameters (λ, θ) , the end results are not analytically-recognized distributions. So, they must be sampled using a different step, such as such as Metropolis sampling, Slice sampling, gradient methods, or gridded approximations (see Wolf et al. (2018) for discussion). We opt for Metropolis in this example; Metropolis-within-Gibbs sampling has been used consistently in spatial analysis (LeSage 1997) and still retains a generally-acceptable balance between efficiency and speed in spatial sampling problems when compared to more recent samplers (Hoffman and Gelman 2014; Wolf et al. 2018).¹¹

A.1 Conditional Posterior for Unit-Level Effects β

This proceeds similarly to the identification of ζ . The form of the marginal effect conditional posterior is:

$$P(\beta | \mathbf{y}, \zeta, \sigma_e^2, \sigma_z^2, \lambda, \theta) \propto L(\mathbf{y} | \beta, \zeta, \sigma_e^2, \lambda) \times P(\beta) \quad (24)$$

Since all of the rest of the parameters' priors are not functions of β , they can be dropped and proportionality still maintained. Fortunately, this allows the β conditional posterior to behave like a typical Bayesian regression posterior update. Letting $\Sigma_y = \Psi(\lambda, \sigma_e^2)$ for convenience, we expand the expression:

$$\propto \exp \left[-\frac{1}{2} (\mathbf{y} - \mathbf{X}\beta - \Delta\zeta)' \Sigma_y^{-1} (\mathbf{y} - \mathbf{X}\beta - \Delta\zeta) \right] \times \exp \left[-\frac{1}{2} (\beta - \mu_\beta)' \Sigma_{\beta 0}^{-1} (\beta - \mu_\beta) \right] \quad (25)$$

$$\propto \exp \left[-\frac{1}{2} \left(-\beta' \mathbf{X}' \Sigma_y^{-1} (\mathbf{y} - \Delta\zeta) - (\mathbf{y} - \Delta\zeta)' \Sigma_y^{-1} \mathbf{X} \beta + \beta' \mathbf{X}' \Sigma_y^{-1} \mathbf{X} \beta - \beta' \Sigma_{\beta 0}^{-1} \mu_\beta - \mu_\beta' \Sigma_{\beta 0}^{-1} \beta + \beta' \Sigma_{\beta 0}^{-1} \beta \right) \right] \quad (26)$$

$$\propto \exp \left[-\frac{1}{2} \left(-\beta' (\mathbf{X}' \Sigma_y^{-1} (\mathbf{y} - \Delta\zeta) + \Sigma_{\beta 0}^{-1} \mu_\beta) - ((\mathbf{y} - \Delta\zeta)' \Sigma_y^{-1} \mathbf{X} + \mu_\beta' \Sigma_y^{-1}) \beta + \beta' (\mathbf{X}' \Sigma_y^{-1} \mathbf{X} + \Sigma_{\beta 0}^{-1}) \beta \right) \right] \quad (27)$$

Following Lindley and Smith (1972), we complete the square using $b' \Sigma_\beta b$, where:

$$b = \left[\mathbf{X}' \Sigma_y^{-1} (\mathbf{y} - \Delta\zeta) + \Sigma_{\beta 0}^{-1} \mu_\beta \right] \quad \Sigma_\beta = \left(\mathbf{X}' \Sigma_y^{-1} \mathbf{X} + \Sigma_{\beta 0}^{-1} \right)^{-1} \quad (28)$$

¹¹Both Metropolis-within-Gibbs and Slice sampling have been implemented in PySAL for these models (Rey and Anselin 2007; Wolf 2018).

This substitution simplifies the expression:

$$\propto \exp \left[-\frac{1}{2} \left(-\beta' b - b' \beta + \beta' \Sigma_\beta^{-1} \beta + b' \Sigma_\beta b - b' \Sigma_\beta b \right) \right] \quad (29)$$

Since $b' \Sigma_\beta b$ is constant with respect to β , the trailing $-b' \Sigma_\beta b$ can be dropped. This allows the expression in Eq. 29 to be factored into the kernel of a p -variate normal distribution of the form:

$$\begin{aligned} & \propto \exp \left[-\frac{1}{2} (\beta - \Sigma_\beta b)' \Sigma_\beta^{-1} (\beta - \Sigma_\beta b) \right] \\ P(\beta | \mathbf{y}, \dots) & \propto \mathcal{N}(\Sigma_\beta b, \Sigma_\beta) \end{aligned} \quad (30)$$

where terms are defined as in Eq. 28. Note that this posterior involves only unit-level parameters since ζ is contained in the conditioning from Eq. 24. The impact of unit-level spatial correlation on β is retained in the posterior variance through $\Sigma_y = \Psi(\lambda) \sigma_e^2$. The region-level spatial correlation affects β only through the ζ term in the posterior mean.

A.2 Conditional Posterior for Unit-Level Scale σ_e^2

The separability of $\Psi(\delta, \sigma_e^2)$ noted above is required for conditional conjugacy in scale parameters. First, define the conditional posterior for the unit-level scale parameter, $P(\sigma_e^2 | \mathbf{y}, \dots)$:

$$P(\sigma_e^2 | \mathbf{y}, \dots) \propto \mathcal{L}(\mathbf{y} | \dots) \times P(\sigma_e^2) \quad (31)$$

That the prior of λ is not involved from the outset is due to the independence of the σ_e^2 and λ priors. The conditional posterior remains a function of σ_e^2 only because of the separability of λ and σ_e^2 in $\Psi(\lambda, \sigma_e^2)$. This becomes clear during the subsequent conjugation. Expanding the distributions:

$$\propto |\Psi(\lambda, \sigma_e^2)|^{-\frac{1}{2}} \exp \left[-\frac{1}{2} \left((\mathbf{y} - \mathbf{X}\beta - \Delta\zeta)' (\Psi(\lambda, \sigma_e^2))^{-1} (\mathbf{y} - \mathbf{X}\beta - \Delta\zeta) \right) \right] \times (\sigma_e^2)^{-a_e-1} \exp \left[\frac{-b_e}{\sigma_e^2} \right] \quad (32)$$

If $\Psi(\lambda, \sigma_e^2)$ were not separable, the determinant and inverse expressions would block conjugation into a univariate full conditional posterior. However, for separable covariance specifications, these expres-

sions can be separated:

$$|\Psi(\lambda, \sigma_e^2)|^{-\frac{1}{2}} = |\Psi(\lambda)|^{-\frac{1}{2}} (\sigma_e^2)^{-\frac{N}{2}} \quad (\Psi(\lambda, \sigma_e^2))^{-1} = \Psi(\lambda)^{-1} \sigma_e^{-2} \quad (33)$$

Therefore, the scale and the spatial covariance component can be separated. This allows for the resulting distribution to have a standard conditionally-conjugate update. Substituting $\eta = \mathbf{y} - \mathbf{X}\beta - \Delta\zeta$:

$$\propto |\Psi(\lambda)|^{-\frac{1}{2}} (\sigma_e^2)^{-\frac{N}{2}} \exp \left[-\frac{1}{2\sigma_e^2} (\eta' \Psi(\lambda)^{-1} \eta) \right] \times (\sigma_e^2)^{-a_e-1} \exp \left[\frac{-b_e}{\sigma_e^2} \right] \quad (34)$$

$$\propto |\Psi(\lambda)|^{-\frac{1}{2}} (\sigma_e^2)^{-\left(\frac{N}{2}+a_e\right)-1} \exp \left[-\frac{1}{\sigma_e^2} \left(\frac{\eta' \Psi(\lambda)^{-1} \eta}{2} + b_e \right) \right] \quad (35)$$

Finally, $|\Psi(\lambda)|^{-\frac{1}{2}}$ is constant with respect to σ_e^2 , so it can be subsumed into the constant of proportionality. Recognizing what is left, the conditional posterior for σ_e^2 is proportional to the following inverse gamma distribution:

$$P(\sigma_e^2 | \dots) \propto IG \left(\frac{N}{2} + a_e, \frac{\eta' \Psi(\lambda)^{-1} \eta}{2} + b_e \right) \quad (36)$$

Notably, $\Psi(\lambda)^{-1}$ remains a term in the updated distribution even though the determinant drops. Later, σ_e^2 will also remain in the updated parameters for λ , linking the two posteriors of the unit-level covariance explicitly.

A.3 Conditional Posterior for Regional Scale σ_z^2

This derivation proceeds similarly to that for the response-level variance term. Since the likelihood is conditioned on ζ , it is constant with respect to σ_z^2 . Thus, the conjugation mixes over the hierarchical prior of ζ and the prior for σ_z^2 :

$$P(\sigma_z^2 | \mathbf{y}, \dots) \propto P(\zeta | \sigma_z^2, \theta) \times P(\sigma_z^2) \quad (37)$$

This leads to the same form of normal $\times$ inverse gamma conjugation shown before, but with different terms:

$$\propto |\Omega(\theta)|^{-\frac{1}{2}} (\sigma_z^2)^{-\frac{I}{2}} \exp \left[-\frac{\zeta' \Omega(\theta)^{-1} \zeta}{2\sigma_z^2} \right] \times (\sigma_z^2)^{-a_z-1} \exp \left[\frac{-b_z}{\sigma_z^2} \right] \quad (38)$$

$$\propto (\sigma_z^2)^{-(\frac{I}{2}+a_z)-1} \exp \left[-\frac{1}{\sigma_z^2} \left(\frac{\zeta' \Omega(\theta)^{-1} \zeta}{2} + b_z \right) \right] \quad (39)$$

$$\propto IG \left(\frac{I}{2} + a_z, \frac{\zeta' \Omega(\theta)^{-1} \zeta}{2} + b_z \right) \quad (40)$$

Again, the conditional posterior contains the spatial correlation component $\Omega(\theta)^{-1}$ in its inverse gamma scale parameter update, and the conditional posterior for θ will also contain σ_z^2 .

A.4 Conditional Posterior for Unit-Level Correlation Parameter λ

For the unit-level spatial correlation parameter, the conditional posterior has the form:

$$P(\lambda | \mathbf{y}, \dots) \propto \mathcal{L}(\mathbf{y} | \dots) \times P(\lambda) \quad (41)$$

The likelihood only depends on λ through the covariance component, $\Psi(\lambda)$. None of the other priors involve $P(\lambda)$ directly, due to the separability of covariance structures and prior independence. However, conjugate priors for λ (or θ) are unlikely. The following distribution is not a recognized distribution for any common choice of $P(\lambda)$ and Ψ :

$$\propto |\Psi(\lambda)|^{-\frac{1}{2}} \exp \left[-\frac{1}{2\sigma_e^2} \left(\eta' \Psi(\lambda)^{-1} \eta \right) \right] \times P(\lambda) \quad (42)$$

While this might appear as a multivariate normal distribution in terms of η , the function is *not* in terms of η . However, this distribution can be sampled efficiently using a different Markov chain Monte Carlo sampling method, like a Metropolis-Hastings sampling step. Since $P(\lambda)$ is constant with respect to other parameters, all of the conditional posteriors derived above are not affected by the choice of $P(\lambda)$. Therefore, any prior on λ may be specified, like an uninformative flat prior, or more informative priors like a scaled-shifted $\mathcal{B}(a_\lambda, b_\lambda)$ or a truncated $\mathcal{N}(\mu_\lambda, \sigma_\lambda^2)$ and a similar (yet likely still analytically unrecognized) conditional posterior would result. In addition, note that the variance term σ_e^2 is retained in the kernel of the conditional posterior, completing the linking mentioned above between λ and σ_e^2 in

their posteriors.

A.5 Conditional Posterior for Regional Correlation Parameter θ

This proceeds similarly to the discussion for λ . Instead of conjugating over the likelihood, however, the conjugation occurs solely in the region-level:

$$P(\theta|\mathbf{y}, \dots) \propto P(\zeta|\theta, \sigma_z^2) \times P(\theta) \quad (43)$$

When expanded, this yields:

$$\propto |\Omega(\theta)|^{-\frac{1}{2}} \exp \left[-\frac{1}{2\sigma_z^2} \left(\zeta' \Omega(\theta)^{-1} \zeta \right) \right] \times P(\theta) \quad (44)$$

This is similar in form to the conditional posterior for λ in Eq. 42, but applies to the region-level model rather than the unit-level model. It must be sampled using a Metropolis step within the Gibbs sampler. It also exhibits an analogous link to the region-level scale, σ_z^2 , as λ is linked to σ_e^2 .

B Appendix: Multilevel spatial lag models require different treatment

For simplicity, we focus on the minimal case, where there is endogenous spatial lag in the varying intercept process, but both region-level and response-level errors are independent and identically distributed:

$$\begin{aligned}
 \mathbf{y} &= \mathbf{X}\beta + \Delta\alpha + \epsilon \\
 \alpha &= \omega\mathbf{M}\alpha + \alpha_{j0} + \mathbf{Z}\gamma + \zeta \\
 \epsilon &\sim \mathcal{N}(0, \sigma_\epsilon^2) \\
 \zeta &\sim \mathcal{N}(0, \sigma_\zeta^2)
 \end{aligned} \tag{45}$$

where $\mathbf{M}$ is a spatial weights matrix describing the geographical structures relating each group to all other groups.

$$\mathbf{y} = \mathbf{X}\beta + \Delta(\mathbf{I} - \omega\mathbf{M})^{-1}\alpha_{j0} + \Delta(\mathbf{I} - \omega\mathbf{M})^{-1}\mathbf{Z}\gamma + \Delta(\mathbf{I} - \omega\mathbf{M})^{-1}\zeta + \epsilon \tag{46}$$

This reduced form yields terms $\mathbf{Z}\gamma$ and $\mathbf{X}\beta$ that cannot be combined the same manner as in the variance components case. Specifically, the interaction between the spatial filter term $(\mathbf{I} - \omega\mathbf{W})^{-1}$ and γ prevents the inclusion of γ and α_{j0} in β^* and $\Delta\mathbf{Z}$ in $\mathbf{X}^*$.

Substantively, this is the same kind of interaction between ρ and β as seen in the classic single-level spatial lag model. In the multilevel model specified in Eq. 46, though, this interaction occurs at the region level. Endogenous inter-regional dependence does not affect any elements in $\mathbf{X}\beta$, the model of the mean of $\mathbf{y}$. However, if reparameterized, β^* would become a function of three parameters (β, γ, ω) where only some of the marginal effects interacted with ω . Further, ω would not be *fully* contained within β^* , since it also interacts with ζ . Most troubling, β^* would no longer be interpretable as a classic marginal effect. In the same manner as LeSage and Pace (2009, §2.7.1) for single-level models, the marginal effect of γ in the region-level model for α (and thus on $\mathbf{y}$) now involves both its “direct” impact and its “indirect” spillovers. Its inclusion in β^* would confuse the interpretation of β alone, which are is affected by region-level dependence. While there still are highly-complex interactions between regional effects in the spatial multilevel variance components model that we discuss later, the marginal effect estimates from spatially-correlated variance components models can be interpreted directly. However,

models like that of Dong and Harris (2015) or the general spatial model of Anselin (1988) are not easily restated into variance components and require a different analysis that accounts for the direct/indirect spillover structure.

DRAFT

C Appendix: Statements of the Shrinkage Matrix

One caveat is useful before proceeding. We use an estimate of λ here because shrinkage in multilevel models is examined post-estimation. However, in some cases $\hat{\lambda}$ from the spatial multilevel specification may be different from a corresponding single-level estimate. For spatial multilevel shrinkage, use the spatial multilevel estimate. We thank an anonymous reviewer for this caveat.

Starting again from the form discussed in section :

$$\mathbf{E}[\hat{\zeta}_{smlm}] = \Sigma_{\zeta} \mathbf{z} \quad (47)$$

$$= \left[\Delta' \mathbf{F}' \mathbf{F} \Delta \sigma_e^{-2} + \Omega(\theta, \sigma_z^2)^{-1} \right]^{-1} \Delta' \mathbf{F}' \mathbf{F} \sigma_e^{-2} \mathbf{y} \quad (48)$$

To acquire a shrinkage matrix like Eq. 5, we will aim for an expression with a single-level estimate of ζ premultiplied by a shrinkage matrix. Ideally, a direct analogue of the result in Jones and Bullen (1994) will also contain one term due explicitly to the variance in upper-level effects, σ_z^2 , and one term due to the remaining "nuisance" variation. An alternative statement with a much lower computational cost is provided in the next section. This alternative form keeps region-level and response-level covariances separate, but is not in the same form as Jones and Bullen (1994).

First, we can aim to recover $\hat{\zeta}_{se}$ for the right term. To get this, we factor out $(\Delta' \mathbf{F}' \mathbf{F} \Delta)$ from the left:

$$= \left[(\Delta' \mathbf{F}' \mathbf{F} \Delta) \left(\mathbf{I}_J \sigma_e^{-2} + (\Delta' \mathbf{F}' \mathbf{F} \Delta)^{-1} \Omega(\theta, \sigma_z^2)^{-1} \right) \right]^{-1} \sigma_e^{-2} \Delta' \mathbf{F}' \mathbf{F} \mathbf{y} \quad (49)$$

Then, recognizing that $(\mathbf{A}\mathbf{B})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$, we can move $(\Delta' \mathbf{F}' \mathbf{F} \Delta)$ to obtain $\hat{\zeta}_{se}$:

$$= \left(\mathbf{I}_J \sigma_e^{-2} + (\Delta' \mathbf{F}' \mathbf{F} \Delta)^{-1} \Omega(\theta, \sigma_z^2)^{-1} \right)^{-1} \sigma_e^{-2} (\Delta' \mathbf{F}' \mathbf{F} \Delta)^{-1} \Delta' \mathbf{F}' \mathbf{F} \mathbf{y} \quad (50)$$

$$= \left(\mathbf{I}_J \sigma_e^{-2} + (\Delta' \mathbf{F}' \mathbf{F} \Delta)^{-1} \Omega(\theta, \sigma_z^2)^{-1} \right)^{-1} \sigma_e^{-2} \hat{\zeta}_{se} \quad (51)$$

Finally, we rearrange the σ terms. To do this, we separate the upper-level covariance into $\Omega(\theta) \sigma_z^2$:

$$= \left(\mathbf{I}_J \sigma_e^{-2} + (\Delta' \mathbf{F}' \mathbf{F} \Delta)^{-1} \Omega(\theta)^{-1} \sigma_z^{-2} \right)^{-1} \sigma_e^{-2} \hat{\zeta}_{se} \quad (52)$$

Finding a common denominator in the expression, like $(\mathbf{C}a^{-1} + \mathbf{D}b^{-1})^{-1} = (\mathbf{C}b + \mathbf{D}a)^{-1}(ab)$, allows

us to further simplify the shrinkage matrix:

$$= \left(\mathbf{I}_J \sigma_z^2 + (\Delta' \mathbf{F}' \mathbf{F} \Delta)^{-1} \Omega(\theta)^{-1} \sigma_e^2 \right)^{-1} (\sigma_z^2 \sigma_e^2) \sigma_e^{-2} \hat{\zeta}_{se} \quad (53)$$

Cancelling extra terms, this results in a shrinkage matrix like Jones and Bullen (1994):

$$\mathbf{E}[\hat{\zeta}_{smlm}] = \left[\left(\mathbf{I}_J \sigma_z^2 + (\Delta' \mathbf{F}' \mathbf{F} \Delta)^{-1} \Omega(\theta)^{-1} \sigma_e^2 \right)^{-1} \sigma_z^2 \right] \hat{\zeta}_{se} \quad (54)$$

The matrix in brackets is the $J \times J$ shrinkage matrix for the spatial multilevel variance components model.

C.1 Standard form of the Shrinkage Matrix

C.2 An alternative statement

Starting again from the statement in Eq. 47, we can obtain a much more numerically-stable statement of the shrinkage matrix if we do not strive for parity with the Jones and Bullen (1994) result. Instead, we let obtain a different relationship between the regional & response-level spatial covariance and the hierarchical covariance. To do this, we can introduce the equivalent of an identity matrix (shown first in braces) and then exploit associativity:

$$\mathbf{E}[\hat{\zeta}_{smlm}] = \left[\Delta' \mathbf{F}' \mathbf{F} \Delta \sigma_e^{-2} + \Omega(\theta, \sigma_z^2)^{-1} \right]^{-1} (\Delta' \mathbf{F}' \mathbf{F}) \sigma_e^{-2} \mathbf{y} \quad (55)$$

$$= \left[\Delta' \mathbf{F}' \mathbf{F} \Delta \sigma_e^{-2} + \Omega(\theta, \sigma_z^2)^{-1} \right]^{-1} \sigma_e^{-2} \left\{ (\Delta' \mathbf{F}' \mathbf{F} \Delta) (\Delta' \mathbf{F}' \mathbf{F} \Delta)^{-1} \right\} (\Delta' \mathbf{F}' \mathbf{F}) \mathbf{y} \quad (56)$$

$$= \left[\Delta' \mathbf{F}' \mathbf{F} \Delta \sigma_e^{-2} + \Omega(\theta, \sigma_z^2)^{-1} \right]^{-1} \sigma_e^{-2} (\Delta' \mathbf{F}' \mathbf{F} \Delta) \left[(\Delta' \mathbf{F}' \mathbf{F} \Delta)^{-1} (\Delta' \mathbf{F}' \mathbf{F}) \mathbf{y} \right] \quad (57)$$

$$= \left[\Delta' \mathbf{F}' \mathbf{F} \Delta \sigma_e^{-2} + \Omega(\theta, \sigma_z^2)^{-1} \right]^{-1} \sigma_e^{-2} (\Delta' \mathbf{F}' \mathbf{F} \Delta) \hat{\zeta}_{se} \quad (58)$$

This result is mathematically-equivalent, but differs slightly in how it relates the “substantive” variation in the effect (σ_z^2) and the autoregressive impacts on the substantive factors. This statement is much more numerically stable for SAR models, though, and should always be used for these cases. This is more stable because there is no intermediate matrix inversion of $\Delta' \mathbf{F}' \mathbf{F} \Delta$ or (assuming Ω is a SAR error correlation as well) $\mathbf{G}' \mathbf{G}$; one matrix is inverted one time. The statement in Eq. 8 involves an additional inversion of $\Delta' \mathbf{F}' \mathbf{F} \Delta$ and increases the impact of floating-point accumulation errors.

D Appendix: Estimate Tables

	SFE OLS	MLM	MLM Upper SE	SFE-SE	MLM Lower SE	MLM Dual SE
AL	0.320 (0.56)	-0.029 (0.14)	-0.135 (0.23)	-0.255 (0.64)	0.032 (0.18)	-0.036 (0.28)
AZ	-0.435 (0.59)	-0.647 (0.23)	-0.805 (0.29)	-1.178 (0.70)	-0.449 (0.29)	-0.693 (0.35)
AR	-0.280 (0.55)	-0.608 (0.13)	-0.732 (0.23)	-0.903 (0.63)	-0.498 (0.18)	-0.649 (0.28)
CA	0.436 (0.55)	0.087 (0.15)	-0.099 (0.24)	-0.500 (0.64)	-0.088 (0.21)	-0.366 (0.31)
CO	-0.069 (0.56)	-0.391 (0.14)	-0.548 (0.23)	-0.624 (0.63)	-0.195 (0.19)	-0.440 (0.29)
CT	0.825 (0.63)	0.346 (0.28)	0.280 (0.32)	0.150 (0.75)	0.181 (0.31)	0.166 (0.37)
DE	1.841 (0.75)	0.799 (0.38)	0.556 (0.39)	0.928 (0.81)	0.476 (0.34)	0.285 (0.39)
FL	1.198 (0.55)	0.817 (0.13)	0.673 (0.23)	0.491 (0.63)	0.641 (0.20)	0.495 (0.29)
GA	0.376 (0.56)	0.027 (0.12)	-0.080 (0.22)	-0.207 (0.63)	0.073 (0.15)	0.005 (0.26)
ID	-0.464 (0.55)	-0.763 (0.15)	-0.878 (0.24)	-1.323 (0.64)	-0.801 (0.21)	-0.912 (0.30)
IL	0.306 (0.56)	-0.054 (0.12)	-0.169 (0.22)	-0.448 (0.63)	-0.129 (0.16)	-0.241 (0.27)
IN	-0.144 (0.57)	-0.486 (0.12)	-0.590 (0.22)	-0.694 (0.64)	-0.315 (0.17)	-0.413 (0.27)
IA	0.609 (0.56)	0.236 (0.12)	0.132 (0.22)	-0.191 (0.64)	0.080 (0.16)	-0.008 (0.27)
KS	0.197 (0.56)	-0.163 (0.12)	-0.291 (0.22)	-0.470 (0.63)	-0.107 (0.16)	-0.297 (0.27)
KY	0.219 (0.54)	-0.127 (0.12)	-0.256 (0.22)	-0.301 (0.60)	0.035 (0.15)	-0.120 (0.27)
LA	0.239 (0.55)	-0.109 (0.14)	-0.252 (0.23)	-0.157 (0.64)	0.165 (0.19)	-0.067 (0.29)
ME	1.251 (0.58)	0.770 (0.22)	0.620 (0.28)	0.595 (0.72)	0.446 (0.32)	0.373 (0.38)
MD	0.981 (0.59)	0.564 (0.19)	0.432 (0.26)	0.287 (0.67)	0.358 (0.22)	0.229 (0.31)
MA	0.397 (0.60)	0.025 (0.24)	-0.025 (0.29)	-0.288 (0.71)	-0.079 (0.28)	-0.039 (0.36)
MI	1.092 (0.55)	0.715 (0.12)	0.576 (0.22)	0.376 (0.63)	0.560 (0.19)	0.395 (0.29)
MN	1.614 (0.56)	1.214 (0.13)	1.104 (0.22)	0.888 (0.63)	1.008 (0.18)	0.932 (0.27)
MS	0.004 (0.56)	-0.334 (0.14)	-0.458 (0.23)	-0.423 (0.63)	-0.084 (0.18)	-0.219 (0.28)
MO	0.276 (0.55)	-0.076 (0.12)	-0.205 (0.22)	-0.447 (0.62)	-0.114 (0.15)	-0.280 (0.27)
MT	0.616 (0.55)	0.255 (0.14)	0.132 (0.23)	-0.179 (0.64)	0.122 (0.20)	0.000 (0.29)
NE	-0.295 (0.57)	-0.638 (0.13)	-0.745 (0.22)	-1.064 (0.64)	-0.660 (0.17)	-0.770 (0.27)

Table 5: State estimates & standard errors for models shown in Figure 1

	SFE OLS	MLM	MLM Upper SE	SFE-SE	MLM Lower SE	MLM Dual SE
NV	-1.086 (0.59)	-1.221 (0.22)	-1.323 (0.28)	-1.430 (0.67)	-0.653 (0.27)	-0.905 (0.34)
NH	0.333 (0.61)	-0.019 (0.25)	0.016 (0.30)	-0.182 (0.73)	0.036 (0.30)	0.102 (0.36)
NJ	0.187 (0.59)	-0.155 (0.20)	-0.231 (0.26)	-0.504 (0.69)	-0.159 (0.26)	-0.190 (0.33)
NM	-0.197 (0.57)	-0.484 (0.18)	-0.649 (0.26)	-0.647 (0.65)	-0.157 (0.22)	-0.445 (0.32)
NY	0.497 (0.55)	0.132 (0.14)	0.007 (0.23)	-0.174 (0.64)	0.101 (0.20)	-0.019 (0.29)
NC	0.726 (0.57)	0.365 (0.12)	0.244 (0.22)	0.141 (0.64)	0.387 (0.16)	0.273 (0.27)
ND	1.343 (0.56)	0.933 (0.15)	0.842 (0.23)	0.672 (0.65)	0.773 (0.21)	0.730 (0.29)
OH	-0.042 (0.56)	-0.388 (0.12)	-0.499 (0.22)	-0.605 (0.63)	-0.239 (0.17)	-0.346 (0.27)
OK	-0.257 (0.56)	-0.590 (0.13)	-0.716 (0.22)	-0.889 (0.64)	-0.469 (0.17)	-0.648 (0.28)
OR	-0.030 (0.56)	-0.354 (0.16)	-0.477 (0.24)	-0.657 (0.65)	-0.229 (0.22)	-0.441 (0.31)
PA	0.029 (0.56)	-0.321 (0.13)	-0.421 (0.23)	-0.514 (0.64)	-0.192 (0.18)	-0.280 (0.28)
RI	1.092 (0.67)	0.475 (0.32)	0.386 (0.35)	0.275 (0.76)	0.236 (0.32)	0.215 (0.39)
SC	0.659 (0.58)	0.292 (0.16)	0.180 (0.24)	0.079 (0.66)	0.284 (0.21)	0.214 (0.29)
SD	1.035 (0.57)	0.651 (0.14)	0.537 (0.23)	0.082 (0.64)	0.297 (0.18)	0.222 (0.28)
TN	-0.036 (0.56)	-0.373 (0.13)	-0.489 (0.22)	-0.506 (0.63)	-0.155 (0.16)	-0.272 (0.27)
TX	-0.138 (0.56)	-0.470 (0.10)	-0.604 (0.21)	-0.990 (0.64)	-0.620 (0.14)	-0.775 (0.26)
UT	0.151 (0.57)	-0.175 (0.17)	-0.386 (0.26)	-0.530 (0.66)	-0.058 (0.23)	-0.401 (0.33)
VT	0.327 (0.59)	-0.023 (0.23)	-0.126 (0.29)	-0.513 (0.70)	-0.159 (0.28)	-0.163 (0.35)
VA	0.433 (0.56)	0.079 (0.12)	-0.040 (0.22)	-0.179 (0.63)	0.099 (0.15)	-0.008 (0.27)
WA	0.872 (0.55)	0.487 (0.16)	0.305 (0.25)	0.096 (0.65)	0.338 (0.23)	0.025 (0.33)
WV	0.130 (0.53)	-0.211 (0.14)	-0.330 (0.23)	-0.504 (0.62)	-0.176 (0.18)	-0.279 (0.28)
WI	0.935 (0.56)	0.555 (0.13)	0.459 (0.22)	0.208 (0.64)	0.404 (0.19)	0.363 (0.28)
WY	-0.344 (0.58)	-0.614 (0.19)	-0.730 (0.26)	-0.868 (0.66)	-0.359 (0.24)	-0.551 (0.32)

Table 6: State estimates & standard errors for models shown in Figure 1