Modeling Hierarchical Spatial Interdependence for Limited Dependent Variables*

Ali Kagalwala† and Hankyeul Yang‡

September 19, 2024

Please do not share or cite without the authors' permission.

Abstract

Scholars are often interested in questions that examine interactions between units at two different levels. For example, counties are nested within states and diffusion processes might take place at both levels of analysis. Building on recent research from the spatial econometrics and multilevel modeling literature, we propose a method for modeling spatial interdependencies in two hierarchical levels with binary and ordered outcomes. We propose a Bayesian approach that estimates spatial autoregressive parameters at both hierarchical levels. Our Monte Carlo results demonstrate that failing to account for the nested structure of the data leads to biased estimates of effects of insterest. We demonstrate the utility of our approach by analyzing the causes of civil rights protests in the United States in the 1960s.

*We thank Scott Cook, Casey Crisman-Cox, Christine Lipsmeyer, Guy Whitten and participants at 2024 TexMeth for their useful comments and suggestions.

†Visiting Assistant Professor, Department of Political Science, Texas A&M University, Email: alikagalwala@tamu.edu.

‡PhD candidate, Department of Political Science, Texas A&M University, Email: yanghankyeul@tamu.edu.

1 Introduction

Analyzing data that have a nested structure is popular in Political Science. Two examples of data that contain a nested structure include units that are each observed across multiple time periods (e.g., time-series-cross-section data), and units at different levels of analysis, (e.g., voters in a county). Due to the hierarchical nature of the data, estimating models that fail to account for the nested structure of the data can lead to inaccurate inferences.

Another popular methodological approach that has been garnering popularity is the spatial analysis of political data. Many Political Science theories involve the diffusion of some policy across units or an occurrence of a certain phenomenon in some units affecting outcomes in other units. For example, Franzese, Hays and Cook (2016) model the civil war diffusion process in Sub-Saharan Africa and estimate the increased probability of civil war in Guinea-Bissau due to internal instability in Senegal. A budding vein of research now combines advances in spatial econometrics and hierarchical modeling to model diffusion at potentially multiple levels of analysis.

We contribute by introducing a spatial-hierarchical strategy for categorical outcomes that are binary or ordered. We introduce a hierarchical spatial probit approach for binary and ordered outcomes.[1] We use a well-known "data augmentation algorithm" to extend the hierarchical spatial model for continuous outcomes to binary and ordered outcomes (Albert and Chib, 1993). The entire process ultimately involves generating simulated values of the latent outcome variable $\mathbf{y}^*$ by using a multivariate truncated normal distribution, and then sequentially drawing from full conditional posterior distributions for other parameters of

[1] Our proposed methodology is easily extendable to cases with ordered outcomes. We discuss this in the appendix.

interest as in the continuous case.

We reach two main conclusions from a series of Monte Carlo experiments. First, as expected, estimating a spatial autoregressive model that accounts for diffusion at only one level results in inaccurate inferences when there is diffusion at multiple levels. Second, we find that our proposed hierarchical spatial model that accounts for diffusion across units at multiple levels can be used as a general modeling strategy. We find that estimating our recommended model results in accurate inferences in such a scenario; the spatial parameter for the level in which diffusion does not occur is statistically insignificant. While intuitive, this is important because the lack of false positives highlights how our model can be used as a robustness check to verify the lack of diffusion among higher-level units.[2]

Spatial Econometrics and Hierarchical Models

For the most part, hierarchical models and spatial econometrics have developed as two distinct fields.[3] The hierarchical modeling literature implicitly assumes that units in a common group share some similar characteristics. The typical example often used is that students in the same classroom may often have correlated errors because they have the same teacher and share similar experiences within the classroom. A common approach in economics has been to model the unobserved relationship through the use of robust standard errors (e.g., Angrist, Bettinger and Kremer, 2006). In contrast, hierarchical models explicitly deal with this through the use of multilevel models by nesting lower-level units within high-level units. One of the main purposes of hierarchical models is to achieve efficiency through partial-pooling

[2]This is assuming the estimator satisfies all assumptions for unbiased parameter estimates and accurate standard errors.

[3]Similar to Gelman et al. (2013), we use the terms multilevel models and hierarchical models interchangeably.

and achieve efficiency by borrowing strength (Jackman, 2009, 307).

The focus of the spatial econometrics literature is to explicitly model theoretically interesting diffusion processes. For example, Simmons and Elkins (2004) examines how liberal economic policies have diffused across countries over time. There have been recent advances in the spatial econometrics literature that focus on combining multilevel and spatial modeling. For example, Dong and Harris (2015) proposed a hierarchical spatial autoregressive model with which they estimated the effects of various factors on the leasing price of land parcels in China, where land parcels are grouped into various districts. Advances in spatial and hierarchical modeling combined with data availability has created opportunities to improve our understanding of politics. For example, while Mazumder (2018) investigates the persistent effects of civil rights protests, scholars might also want to investigate the conditions under which protests are likely to emerge.

2 Spatial Autoregressive Model

One of the most popular models used in spatial econometrics is the spatial autoregressive (SAR) model, [4] which can be expressed as

$$\mathbf{y} = \rho \mathbf{W} \mathbf{y} + \mathbf{X} \boldsymbol{\beta} + \boldsymbol{\epsilon}, \tag{1}$$

where $\mathbf{y}$ is an $N \times 1$ vector of outcomes in all N units, ρ is the spatial autoregressive coefficientm $\mathbf{X}$ is a matrix of predictors with dimensions $N \times K$, $\boldsymbol{\beta}$ is an $K \times 1$ vector, and $\mathbf{W}$ is an $N \times N$ spatial weights matrix.

[4] Readers interested in the full range of commonly models used in spatial econometrics can refer to Halleck Vega and Elhorst (2015).

This model has been used across a broad range of applications that seek to understand how the outcome in one unit affects those in others, such as party positions (Williams and Whitten, 2015) and military spending (George and Sandler, 2018). However, there are cases in which scholars may be interested in modeling the relationship between outcomes of higher-level units in addition to the diffusion in outcomes among lower-level units. This is similar to the motivation behind the use of hierarchical models, but a little different in terms of its focus. For example, hierarchical models might be used to study the effect of voter characteristics on which party they vote for. Naturally, voters are grouped into different states and the errors amongst the voters in a given state might be correlated. One advantage of hierarchical modeling has been to increase efficiency by partial pooling (Ghitza and Gelman, 2013).[5]

3 Hierarchical Spatial Probit Autoregressive Model

We propose a binary outcome model for estimating spatial autoregressive coefficients that accounts for the hierarchical nature of data. Our proposed hierarchical spatial autoregressive (HSAR) probit model is an extension of the work of Dong and Harris (2015), who focused on the continuous outcomes case.[6]

We can first conceptualize the outcome as a continuous latent variable, $\mathbf{y}^*$, such that

$$\mathbf{y}^* = \rho \mathbf{W} \mathbf{y}^* + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\Delta}\boldsymbol{\theta} + \boldsymbol{\epsilon}, \tag{2}$$

[5] Partial pooling may be thought of as “compromising between the two extremes of excluding a categorical predictor from a model (complete pooling), or estimating separate models within each level of the categorical predictor (no pooling).” (Gelman and Hill, 2006, 252)

[6] Without loss of generality, our model is a stylized version in which higher-level predictors have been omitted. As we show later, this is to make our model easily comparable to the original SAR probit model. Practitioners can easily extend the model presented here to include covariates for the higher-level units.

where there are $i = 1, ..., N$ lower-level units nested in $j = 1, ..., J$ higher-level units. $\mathbf{y}^*$ is a $N \times 1$ vector of the latent, continuous outcome variable, $\mathbf{W}$ is an $N \times N$ spatial weights matrix for lower-level units, $\mathbf{X}$ is an $N \times K$ matrix of covariates, $\boldsymbol{\beta}$ is a $K \times 1$ vector of parameters, $\boldsymbol{\Delta}$ is a $N \times J$ matrix mapping lower-level units to higher-level units, $\boldsymbol{\theta}$ is a $J \times 1$ vector of random effects, and $\boldsymbol{\epsilon}$ is an $N \times 1$ vector of white noise.[7]

The model assumes that $\boldsymbol{\theta}$ follows its own autoregressive process and that there is a spatial weights matrix $\mathbf{M}$ of dimensions $J \times J$ as shown below

$$\boldsymbol{\theta} = \lambda \mathbf{M} \boldsymbol{\theta} + \mathbf{u} \Rightarrow \boldsymbol{\theta} = (\mathbf{I} - \lambda \mathbf{M})^{-1} \mathbf{u} \tag{4}$$

$$\mathbf{u} \sim \mathcal{N}(\mathbf{0}, \sigma_u^2 \mathbf{I}_J) \tag{5}$$

$$\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_N) \tag{6}$$

$$\boldsymbol{\theta} \sim \mathcal{N}(0, \sigma_u^2 (\mathbf{B}' \mathbf{B})^{-1}) \text{ where } \mathbf{B} \equiv \mathbf{I}_J - \lambda \mathbf{M} \tag{7}$$

As is standard in the literature, we assume that we observe a value of 1 for the binary outcome, y_{ij}, if the latent variable y_{ij}^* is (weakly) positive and 0 otherwise:

$$y_{ij} = 1 \Longleftrightarrow y_{ij}^* \geq 0 \tag{8}$$

$$y_{ij} = 0 \Longleftrightarrow y_{ij}^* < 0 \tag{9}$$

[7] As an example of a $\boldsymbol{\Delta}$ matrix, consider a 6×3 $\boldsymbol{\Delta}$ matrix consisting of 6 lower-level units with 2 lower-level units in each higher-level unit. This can be represented by the following $\boldsymbol{\Delta}$ matrix.

$$\begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix} \tag{3}$$

Dong and Harris (2015) estimate the parameters of interest by adopting a Bayesian approach and implementing Markov Chain Monte Carlo (MCMC) methods. We use a similar approach for our hierarchical spatial models for discrete outcomes. In particular, we use the standard Metropolis-within-Gibbs algorithm in which the parameters are estimated using a Gibbs algorithm whenever the full conditional distributions are of a known form and with a Metropolis algorithm when the distribution is not of a recognizable form. This is standard in the Bayesian literature and has also been used in spatial econometrics to estimate the standard SAR model (LeSage and Pace, 2009).[8]

Estimating a binary outcome model using probit is a simple extension of estimating a linear model under the Bayesian framework. If we *assume* that we observe y^*, we can proceed as usual in estimating the relevant parameters of interest. This is the insight that Albert and Chib (1993) provide in estimating the probit model and has been referred in the Bayesian literature as data augmentation based on Tanner and Wong (1987). Of course, since we do not actually observe y^*, we need to generate them. Intuitively, the appropriate values of y^* can be generated under the restriction that y_i^* is positive (negative) if the observed y_i is 1 (0). Conceptually, this involves using a truncated normal distribution which can be implemented through other more efficient sampling algorithms such as inversion sampling.[9] It is well-known that the rejection sampling can suffer from computational inefficiencies

[8]The Metropolis-within-Gibbs sampling technique has been the standard algorithm for estimating spatial econometric models when using Bayesian methods. We are aware that Bayesian statistics has seen a development of various Markov Chain Monte Carlo algorithms. In particular, the Hamiltonian Monte Carlo algorithm popularized with the development of Stan (Carpenter et al., 2017) has seen wide usage for estimating various models due to its flexibility and ease of use. However, we found that Stan is too computationally inefficient for estimating spatial econometric models with a large number of observations. Wolf, Anselin and Arribas-Bel (2018) suggest that this is due to the computational burden of calculating the log determinant term for each leapfrog step. We refer the readers to Wolf, Anselin and Arribas-Bel (2018) for more details on this matter.

[9]An inversion sampling mechanism involving using a draw from a uniform distribution and the inverse-distribution function (Lynch, 2007, 203).

(Lynch, 2007) and often a form of inversion sampling is implemented in practice.

However, there is a crucial difference when estimating the probit model in the spatial econometrics literature. In contrast to the standard probit model, the errors are no longer assumed to be independent and the error structure is more complex. We need to thus use truncated multivariate normal distribution to generate the latent variables $\mathbf{y}^*$ instead of (independent) truncated univariate normal distributions as is the case with the standard probit model (e.g., Geweke, 1991). Standard results in statistics demonstrate that conditional distributions of a truncated multivariate normal distribution can be reduced to a truncated univariate normal distribution (Geweke, 1991; Kotecha and Djuric, 1999). Thus, a type of Gibbs sampling procedure can be implemented to sample from a truncated multivariate normal distribution.[10]

The likelihood can be represented as follows for the probit model with binary outcomes :(Jackman, 2009)

$$\mathcal{L}(\beta; y, X) = p(y|X, \beta) = \prod_{i=1}^{n} F(x_i\beta)^{y_i}[1 - F(x_i\beta)^{1-y_i}] \tag{10}$$

where $F(\cdot)$ represents the cumulative normal distribution.

Deriving the posterior distribution for the probit model in practice requires the use of the data augmentation approach for probit models (Tanner and Wong, 1987; Albert and Chib, 1993). This essentially involves estimating a new set of parameters, namely the set of N latent variables (e.g., Smith and LeSage, 2004).

We take a similar approach by adapting the derivations from Dong and Harris (2015).

[10] In practice, this can be implemented conveniently using the package **tmvtnorm** in **R** (Wilhelm and Manjunath, 2010).

Our first assumption is that the variance of the error, σ_e^2, is 1 for identification purposes. Second, we assume that there is no covariance between $\boldsymbol{\epsilon}$ and $\boldsymbol{\theta}$ (the vector of random effects). These can be written as:

$$Var(\boldsymbol{\epsilon}) = \mathbf{I}_N \tag{11}$$

$$Cov(\boldsymbol{\epsilon}, \boldsymbol{\theta}) = 0 \tag{12}$$

Based on these above assumptions, the variance-covariance matrix of $\mathbf{y}^*$ in Dong and Harris (2015):

$$Var(\mathbf{y}^*) = (\mathbf{I}_N - \rho\mathbf{W})^{-1} Var(\boldsymbol{\Delta\theta} + \boldsymbol{\epsilon})((\mathbf{I}_N - \rho\mathbf{W})^{-1})' \tag{13}$$

$$= \mathbf{A}^{-1} Var(\boldsymbol{\Delta\theta} + \boldsymbol{\epsilon})(\mathbf{A}^{-1})', \text{ where } \mathbf{A} \equiv \mathbf{I}_N - \rho\mathbf{W} \tag{14}$$

$$= \mathbf{A}^{-1}\Big(\boldsymbol{\Delta} Var(\boldsymbol{\theta})\boldsymbol{\Delta}' + Var(\boldsymbol{\epsilon}) + 2Cov(\boldsymbol{\Delta\theta}, \boldsymbol{\epsilon})\Big)(\mathbf{A}^{-1})' \tag{15}$$

$$= \mathbf{A}^{-1}\Big(\boldsymbol{\Delta} Var(\boldsymbol{\theta})\boldsymbol{\Delta}' + Var(\boldsymbol{\epsilon})\Big)(\mathbf{A}^{-1})' \quad \because \quad \text{Equation 12} \tag{16}$$

$$= \mathbf{A}^{-1}\Big(\boldsymbol{\Delta}(\mathbf{B}'\mathbf{B})^{-1}\boldsymbol{\Delta}' + \mathbf{I}_N\Big)(\mathbf{A}^{-1})' \equiv \mathbf{V} \tag{17}$$

where, $\mathbf{B} \equiv \mathbf{I}_J - \lambda\mathbf{M}$ as per equation 7. As will be seen later, $\mathbf{V}$ will play an important role in using the truncated multivariate normal distributions for generating $\mathbf{y}^*$.[11]

[11] In equations 13 and 15, we use the rule $Cov(\mathbf{Ax}) = \mathbf{AxA}'$.

4 Estimating the Model

As previously mentioned, we adopt a Bayesian Markov Chain Monte Carlo (MCMC) method for model estimation. The basic Bayesian identity used for estimating unknown parameters is (Gelman et al., 2013):

$$p(\mathbf{\Theta}|\mathbf{Y}) \propto \mathcal{L}(\mathbf{Y}|\mathbf{\Theta}) \times \pi(\mathbf{\Theta}) \tag{18}$$

$$\text{posterior} \propto \text{likelihood} \times \text{prior} \tag{19}$$

We employ diffuse priors to allow the likelihood to be dominant in estimating the parameters.[12] The priors may be specified as follows:

$$\pi(\boldsymbol{\beta}) \sim \mathcal{N}(\mathbf{c}_0, \mathbf{T}_0) \tag{20}$$

$$\pi(\rho) \sim \mathcal{U}\left[\frac{1}{\nu_{\min}}, 1\right] \propto 1 \tag{21}$$

$$\pi(\lambda) \sim \mathcal{U}\left[\frac{1}{\nu^*_{\min}}, 1\right] \propto 1 \tag{22}$$

$$\pi(\boldsymbol{\theta}|\lambda) \sim \mathcal{N}\Big(\mathbf{0}, ((\mathbf{I} - \lambda\mathbf{M})'(\mathbf{I} - \lambda\mathbf{M}))^{-1}\Big), \tag{23}$$

where $\mathbf{A} = \mathbf{I}_N - \rho\mathbf{W}$, $\nu_{\min}$ is the minimum eigenvalue of the spatial weights matrix for the lower-level units, $\mathbf{W}$, and $\nu^*_{\min}$ is the minimum eigenvalue of the spatial weights matrix of high-level units, $\mathbf{M}$. We employ diffuse priors to let the data dominate the posterior. This approach is standard and has been adopted by spatial econometrics researchers in past works

[12] The basic derivations for the continuous outcome are detailed in Dong and Harris (2015). We reproduce the derivations here with some minor changes such as working with the latent variable y^* instead of y and constraining σ_e^2 to 1 necessary for identification purposes.

(LeSage and Pace, 2009). In practice, this involves centering the prior for $\boldsymbol{\beta}$ on a vector of zero's for the mean $(\mathbf{c}_0)$ with a large variance $(\mathbf{T}_0)$.

The likelihood function in terms of the latent variable $\mathbf{y}^*$ may be specified as follows:

$$\mathcal{L}(\mathbf{y}^*|\rho,\lambda,\boldsymbol{\beta},\boldsymbol{\theta},\sigma_u^2) = (2\pi)^{-N/2}|\mathbf{A}|\exp\left\{-\frac{1}{2}(\mathbf{A}\mathbf{y}^* - \mathbf{X}\boldsymbol{\beta} - \boldsymbol{\Delta\theta})'(\mathbf{A}\mathbf{y}^* - \mathbf{X}\boldsymbol{\beta} - \boldsymbol{\Delta\theta})\right\} \quad (24)$$

The respective conditional distributions for each parameter or set of parameters based on the following Bayesian identity (Dong and Harris, 2015):

$$p(\rho,\lambda,\boldsymbol{\beta},\boldsymbol{\theta},\mathbf{y}^*|\mathbf{y}) \propto \mathcal{L}(\mathbf{y}^*|\rho,\lambda,\boldsymbol{\beta},\boldsymbol{\theta},\sigma_u^2)\cdot\pi(\rho)\cdot\pi(\lambda)\cdot\pi(\boldsymbol{\beta})\cdot\pi(\boldsymbol{\theta}|\lambda,\sigma_u^2)\cdot\pi(\sigma_u^2) \quad (25)$$

$$\text{posterior} \propto \text{likelihood} \times \text{prior} \quad (26)$$

where $\pi(\rho)\cdot\pi(\lambda)\cdot\pi(\boldsymbol{\beta})\cdot\pi(\boldsymbol{\theta}|\lambda,\sigma_u^2)\cdot\pi(\sigma_u^2)$ denote the priors for the respective parameters.

The basic strategy is to employ a combination of Gibbs sampling and the Metropolis-Hasting sampling algorithms.[13] The researcher needs to be able to derive the full conditional distribution in a recognized form to be able to employ the Gibbs sampling algorithm (Smith and LeSage, 2004). As will be seen, while some sets of parameters can be estimated using the Gibbs sampling algorithm, others will have to be estimated using the Metropolis-Hastings algorithm. An additional complication in estimating ρ and λ for discrete outcomes is necessarily having to data augmentation. Below, we reproduce the results of the derivations for the conditional distributions for the probit model by constraining σ_e^2 to 1.

[13]Readers may refer to Jackman (2009) and Gelman et al. (2013) for a detailed reference on these algorithms.

Generating $\mathbf{y}^*$

We generate samples of $\mathbf{y}^*$ as follows (Albert and Chib, 1993):

$$y_{ij}^* | \rho, \boldsymbol{\beta}, \boldsymbol{\theta}, \mathbf{y} \sim \begin{cases} \mathcal{MVN}(k_i, v_i) \mathbb{1}(y_{ij}^* \geq 0) & y_{ij} = 1 \\ \\ \mathcal{MVN}(k_i, v_i) \mathbb{1}(y_{ij}^* < 0) & y_{ij} = 0 \end{cases} \tag{27}$$

where $\mathbb{1}$ is the indicator function, v_i denotes the $\mathbf{V}_{ii}$ element in the variance-covariance matrix of $\mathbf{y}^*$, and k_i is the i^{th} element of the $N \times 1$ column vector $\mathbf{K} \equiv \mathbf{A}^{-1}(\mathbf{X}\boldsymbol{\beta})$. Thus, equation 38 allows us to generate the latent values, y_{ij}^*, while simultaneously accounting for spatially correlated errors. Readers familiar with the data augmentation technique will note that the results above are slightly different from the case of truncated univariate normal distributions as was used for the (non-spatial) probit model by Albert and Chib (1993). The added complication in generating y_{ij}^* here in the context of spatial econometrics is the interdependence of errors (Franzese, Hays and Cook, 2016). Naively using independent truncated univariate normal distributions leads to erroneous inferences in this case. The solution to this problem is to use the truncated multivariate normal distribution (LeSage and Pace, 2009; Wilhelm and de Matos, 2013).

Conditional Posterior Distribution for $\boldsymbol{\beta}$

$$p(\boldsymbol{\beta}|\mathbf{y}^*, \rho, \lambda, \boldsymbol{\theta}, \sigma_u^2) \propto \mathcal{L}(\mathbf{y}^*|\rho, \lambda, \boldsymbol{\beta}, \boldsymbol{\theta}, \sigma_u^2) \cdot \pi(\boldsymbol{\beta}) \tag{28}$$

$$\propto \exp\left\{-\frac{1}{2}(\mathbf{A}\mathbf{y}^* - \mathbf{X}\boldsymbol{\beta} - \boldsymbol{\Delta}\boldsymbol{\theta})'(\mathbf{A}\mathbf{y}^* - \mathbf{X}\boldsymbol{\beta} - \boldsymbol{\Delta}\boldsymbol{\theta})\right\} \times \exp\left\{-\frac{1}{2}(\boldsymbol{\beta} - \mathbf{M}_0)'\mathbf{T}_0^{-1}(\boldsymbol{\beta} - \mathbf{M}_0)\right\} \tag{29}$$

$$\propto \exp\left\{-\frac{1}{2}\boldsymbol{\beta}'[\mathbf{X}'\mathbf{X} + \mathbf{T}_0^{-1}]\boldsymbol{\beta} + [(\mathbf{A}\mathbf{y}^* - \boldsymbol{\Delta}\boldsymbol{\theta})'\mathbf{X} + \mathbf{T}_0^{-1}\mathbf{M}_0]\boldsymbol{\beta} + \mathbf{C}\right\} \tag{30}$$

The basic logic behind deriving the full conditional distributions is to treat priors that do not contain the parameters of interest as constants. This allows us to simplify the original expression summarizing the relationship between the posterior on the one hand and the likelihood and the prior on the other as shown above. We can then just work with the kernel of the distributions by omitting any constants that do not affect the proportionality: equations 29 and 30. show how the conditional distribution for $\boldsymbol{\beta}$ is derived after omitting the priors for the parameters that are not of interest and working with the kernels. It is well-known that this form can be simplified further by making use of the properties of the normal distribution (Smith and LeSage, 2004). If we focus on the terms within the brackets, the expression above simplifies to the following:

$$p(\boldsymbol{\beta}|\mathbf{y}^*, \rho, \lambda, \boldsymbol{\theta}, \sigma_u^2) \sim \mathcal{MVN}(\mathbf{M}_\beta, \boldsymbol{\Sigma}_\beta) \tag{31}$$

where $\boldsymbol{\Sigma}_\beta \equiv [\mathbf{X}'\mathbf{X} + \mathbf{T}_0^{-1}]^{-1}$ and $\mathbf{M}_\beta \equiv \boldsymbol{\Sigma}_\beta[\mathbf{X}'(\mathbf{A}\mathbf{y}^* - \boldsymbol{\Delta}\boldsymbol{\theta}) + \mathbf{T}_0^{-1}\mathbf{M}_0]$. Equation 31 is the simplified expression from combining the likelihood with the prior shown in equation (20). Equation 31 shows that we can use multivariate normal distribution with mean $\mathbf{\Sigma}_\beta$ and variance $\mathbf{M}_\beta$ to draw updated values of $\boldsymbol{\beta}$.

Conditional Posterior Distribution for $\boldsymbol{\theta}$

The logic for updating $\boldsymbol{\theta}$ is very similar to the logic for updating $\boldsymbol{\beta}$ (Dong and Harris, 2015):

$$\pi(\boldsymbol{\theta}|\lambda, \sigma_u^2) \cdot \mathcal{L}(\mathbf{y}^*|\rho, \lambda, \boldsymbol{\beta}, \boldsymbol{\theta}, \sigma_u^2) \tag{32}$$

$$\boldsymbol{\theta}|\lambda, \sigma_u^2 \sim \mathcal{MVN}(\mathbf{M}_\theta, \mathbf{\Sigma}_\theta) \tag{33}$$

where $\mathbf{\Sigma}_\theta \equiv [\mathbf{\Delta}'\mathbf{\Delta} + (\sigma_u^2)^{-1}\mathbf{B}'\mathbf{B}]^{-1}$ and $\mathbf{M}_\theta \equiv \mathbf{\Sigma}_\theta[\mathbf{\Delta}'(\mathbf{AY} - \mathbf{X}\boldsymbol{\beta})]$. Equation 32 shows the full conditional distribution for $\boldsymbol{\theta}$ from which new values may be drawn based on the updated values of λ and σ_u^2.

Conditional Posterior Distribution for $\boldsymbol{\sigma_u^2}$

The conditional posterior distribution for σ_u^2 is (Dong and Harris, 2015):

$$\sigma_u^2 \sim \mathcal{IV}\left(\frac{J}{2} + a_0, \frac{\boldsymbol{\theta}'\mathbf{B}'\mathbf{B}\boldsymbol{\theta}}{2} + b_0\right) \tag{34}$$

where $\mathcal{IV}$ denotes the inverse-gamma distribution and a_0 and b_0 are parameters set a priori by the researcher.[14]

[14] We set these to 0.01 similar to Dong and Harris (2015).

Conditional Posterior Distribution for λ

The conditional posterior distribution for λ is (Dong and Harris, 2015):

$$p(\lambda|\mathbf{y}^*, \rho, \boldsymbol{\beta}, \sigma_u^2, \boldsymbol{\theta}) \propto \pi(\boldsymbol{\theta}|\lambda, \sigma_u^2) \cdot \pi(\lambda) \tag{35}$$

$$\propto |\mathbf{I}_J - \lambda\mathbf{M}| \times \exp\left\{ -\frac{1}{2\sigma_u^2}\boldsymbol{\theta}'\mathbf{B}'\mathbf{B}\boldsymbol{\theta}\right\} \tag{36}$$

where $\mathbf{B} = \mathbf{I}_J - \lambda\mathbf{M}$. Equation 35 is not a distribution of a known form and we will have to use a Metropolis-Hastings sampling algorithm. Instead of working with the acceptance ratio directly, we use a logged-transformed version of the ratio for numerical stability purposes (Hoff, 2009).

Conditional Posterior Distribution for ρ

$$p(\rho|\lambda, \boldsymbol{\beta}, \boldsymbol{\theta}, \mathbf{y}^*, \mathbf{y}) = \frac{p(\rho, \lambda, \boldsymbol{\beta}, \boldsymbol{\theta}, \mathbf{y}^*|\mathbf{y})}{p(\lambda, \boldsymbol{\beta}, \boldsymbol{\theta}, \mathbf{y}^*|\mathbf{y})} \tag{37}$$

$$\propto p(\rho, \lambda, \boldsymbol{\beta}, \boldsymbol{\theta}, \mathbf{y}^*|\mathbf{y}) \tag{38}$$

$$\propto \pi(\rho) \cdot \pi(\mathbf{y}^*|\rho, \lambda, \boldsymbol{\beta}, \boldsymbol{\theta}, \sigma_u^2) \tag{39}$$

$$\propto \det|\mathbf{I}_N - \rho\mathbf{W}| \times \exp\left\{ -\frac{1}{2}\Big(\mathbf{A}\mathbf{y}^* - \mathbf{X}\boldsymbol{\beta} - \boldsymbol{\Delta\theta})\Big)'\Big(\mathbf{A}\mathbf{y}^* - \mathbf{X}\boldsymbol{\beta} - \boldsymbol{\Delta\theta})\Big)\right\}, \tag{40}$$

where $\mathbf{A} = \mathbf{I} - \rho\mathbf{W}$. Similar to the conditional distribution of λ, equation (39) is not a distribution of a known form and we will have to use the Metropolis-Hastings sampling algorithm. Once again, we do not work with the acceptance ratio directly but instead use a logged-transformed version for the purposes of numerical stability (Hoff, 2009).

5 Model Implications

Before we delve into the Monte Carlo simulations, it is worth elaborating what our model entails. First, we note that a multilevel random intercept probit model is a special case of our model when their is no spatial interdependence at both levels (i.e., with the restrictions $\rho = 0$ and $\lambda = 0$).

$$\mathbf{y}^* = \rho\mathbf{W}\mathbf{y}^* + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\Delta}\boldsymbol{\theta} + \boldsymbol{\epsilon} \quad \boldsymbol{\theta} = (\mathbf{I} - \lambda\mathbf{M})^{-1}\mathbf{u} \tag{41}$$

$$\Rightarrow \mathbf{y}^* = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\Delta}\mathbf{u} + \boldsymbol{\epsilon} \quad \text{where} \quad \mathbf{u} \sim \mathcal{N}(\mathbf{0}, \sigma_u^2\mathbf{I}_J) \tag{42}$$

$$\tag{43}$$

The main implication here is that the multilevel random intercept probit model is likely to perform similarly to our proposed HSAR probit models when both λ and ρ are relatively low. However, we would expect the multilevel random intercept probit model to render biased estimates of $\boldsymbol{\beta}$ as ρ increases because the multilevel random intercept probit model is not properly accounting for the spatial diffusion process.

Second, when just $\lambda = 0$, note that our model simplifies into a SAR probit model with a random intercept. It is worth highlighting that this is different from the traditional SAR probit model in the spatial econometrics literature which does not contain a random intercept.

$$\mathbf{y}^* = \rho\mathbf{W}\mathbf{y}^* + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\Delta}\boldsymbol{\theta} + \boldsymbol{\epsilon} \quad \boldsymbol{\theta} = (\mathbf{I}_J - \lambda\mathbf{M})^{-1}\mathbf{u} \tag{44}$$

$$\Rightarrow \mathbf{y}^* = \rho\mathbf{W}\mathbf{y}^* + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\Delta}\mathbf{u} + \boldsymbol{\epsilon} \quad \text{where} \quad \mathbf{u} \sim \mathcal{N}(\mathbf{0}, \sigma_u^2\mathbf{I}_J) \tag{45}$$

Note that even when $\lambda = 0$, using the traditional SAR probit model will still induce bias in estimating $\boldsymbol{\beta}$ because there is a random intercept as noted above that the regular SAR probit model does not account for.[15] In this case, omitting a confounder will induce an *attenuation* bias for the independent variable of interest (Neuhaus and Jewell, 1993; Cramer, 2003). As such, although the random intercept induced by setting $\lambda = 0$ will not be correlated with the independent variable, not accounting for this random effect by using the traditional SAR probit model will still induce attenuation bias in estimating $\boldsymbol{\beta}$ and subsequently, an inflationary bias in estimating ρ.

If the researcher is uncertain about the existence of spatial processes, the cost of estimating our proposed HSAR solution when there are no spatial processes at both levels is efficiency losses. The benefit of treating the HSAR model as a general model is unbiasedness (assuming all other model assumptions hold) regardless of the existence of no spatial process at the both levels, a spatial process at one level, or spatial interdependence at both levels.

[15] At this point, it may worth reminding ourselves that the condition for inducing bias in the parameter for the independent variable of interest for binary outcomes is different from that of a linear additive model. Let us posit a linear model of the form

$$y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \epsilon_i \tag{46}$$

where X_1 is the main independent variable of interest and X_2 is the confounder. When X_2 is correlated with both X_1 and with y, we will obtain a biased estimate for β_1 if we mistakenly omit X_2. The condition of being correlated with both the independent variable variable and the dependent variable is no longer necessary to induce bias in the case of binary outcomes, however, as shown by various scholars (Neuhaus and Jewell, 1993; Cramer, 2003).

Thus, the researcher should carefully weigh the benefits and costs, when deciding on the type of multilevel model to estimate.

6 Calculation of Effects

Researchers on spatial econometrics have warned that we cannot directly infer effects from estimates (e.g., Franzese, Hays and Cook, 2016). Since the dependent variable is discrete, we would be interested in finding out, for example, how the propensity for the ith observation to experience an event would change given that there is some change in the variable x_k for unit i.[16] While the calculation of such effects are not simple, past scholars have shown the derivation for such quantities of interest (Beron and Vijverberg, 2004; Franzese, Hays and Cook, 2016). The derivation below is quite similar to these past derivations.[17]

In the case of the hierarchical spatial probit model, the probability of the ith observation experiencing an event is

$$p(y_i = 1|\mathbf{X}) = p([(\mathbf{I} - \rho\mathbf{W})^{-1}\mathbf{X}\boldsymbol{\beta}]_i + [(\mathbf{I} - \rho\mathbf{W})^{-1}\boldsymbol{\Delta\theta}]_i + [(\mathbf{I} - \rho\mathbf{W})^{-1}\boldsymbol{\epsilon}]_i) \tag{47}$$

$$= p([(\mathbf{I} - \rho\mathbf{W})^{-1}\mathbf{X}\boldsymbol{\beta}]_i + [(\mathbf{I} - \rho\mathbf{W})^{-1}\boldsymbol{\Delta}\mathbf{B}^{-1}\mathbf{u}]_i + [(\mathbf{I} - \rho\mathbf{W})^{-1}\boldsymbol{\epsilon}]_i) \tag{48}$$

$$= p(v_i < [(\mathbf{I} - \rho\mathbf{W})^{-1}\mathbf{X}\boldsymbol{\beta}]_i) \tag{49}$$

$$= \Phi\left\{[(\mathbf{I} - \rho\mathbf{W})^{-1}\mathbf{X}\boldsymbol{\beta}]_i / \sigma_{v_i}\right\} \tag{50}$$

where $\mathbf{B} = \mathbf{I}_J - \lambda\mathbf{M}$, Φ is the CDF of the standard normal distribution and σ_{v_i} is the iith

[16] It is worth highlighting that other quantities of interest can also be calculated in spatial econometrics. For example, a researcher might be interested in how the propensity for the ith observation to experience an event would change given that there is some change in the variable x_k for unit j.

[17] Readers may want to compare the derivations below to those of Franzese, Hays and Cook (2016, 155,160).

element of the variance-covariance matrix $\mathbf{V} \equiv \mathbf{A}^{-1}\Big(\boldsymbol{\Delta}(\mathbf{B}'\mathbf{B})^{-1}\boldsymbol{\Delta}' + \mathbf{I}_N\Big)(\mathbf{A}^{-1})'$ mentioned above.[18]

Let x_{ik} stand for the value of variable x_k for observation i. To calculate the change in propensity for the ith observation to experience an event due to a change in x_{ik}, we apply the chain rule and derive the following

$$\frac{\partial p(y_i = 1|\mathbf{X}, \mathbf{M}, \mathbf{W})}{\partial x_{ik}} = \phi_{\text{pdf}}\Big\{[(\mathbf{I} - \rho\mathbf{W})^{-1}\mathbf{X}\boldsymbol{\beta}]_i / \sigma_{v_i}\Big\}[(\mathbf{I} - \rho\mathbf{W})^{-1}\beta_k]_{ii} / \sigma_{v_i} \qquad (52)$$

7 Monte Carlo Simulations

We conduct a series of Monte Carlo simulations to assess the validity of our proposed HSAR model. We test the robustness of our model across a wide range of parameter combinations and conditions. We set the number of contiguous higher-level units such that $J = \{16, 49\}$. We generate 20 random districts within each J. This results in a combined total of $N = \{320, 980\}$ low-level units. The $N \times J$ matrix $\boldsymbol{\Delta}$ maps each of the lower level units, i, to the higher-level units, j. The spatial weights matrices $\mathbf{W}$ and $\mathbf{M}$ were generated accordingly.

[18]Note that

$$V[(\mathbf{I} - \rho\mathbf{W})^{-1}\boldsymbol{\Delta}\mathbf{B}^{-1}\mathbf{u} + (\mathbf{I} - \rho\mathbf{W})^{-1}\boldsymbol{\epsilon}] = V[(\mathbf{I} - \rho\mathbf{W})^{-1}(\boldsymbol{\Delta}\mathbf{B}^{-1}\mathbf{u} + \boldsymbol{\epsilon})] \qquad (51)$$

renders the variance-covariance matrix $Var(\mathbf{y}^*) = \mathbf{A}^{-1}\Big(\boldsymbol{\Delta}(\mathbf{B}'\mathbf{B})^{-1}\boldsymbol{\Delta}' + \mathbf{I}_N\Big)(\mathbf{A}^{-1})' \equiv \mathbf{V}$ we previously worked out before.

$$\mathbf{y}^* = \rho \mathbf{W} \mathbf{y}^* + \mathbf{X} \boldsymbol{\beta} + \boldsymbol{\Delta} \boldsymbol{\theta} + \boldsymbol{\epsilon} \tag{53}$$

$$\boldsymbol{\theta} = \lambda \mathbf{M} \boldsymbol{\theta} + \mathbf{u} \Rightarrow \boldsymbol{\theta} = (\mathbf{I} - \lambda \mathbf{M})^{-1} \mathbf{u} \tag{54}$$

$$\mathbf{u} \sim \mathcal{N}(\mathbf{0}, \sigma_u^2 \mathbf{I}_J) \tag{55}$$

$$\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_N) \tag{56}$$

$$\boldsymbol{\theta} \sim \mathcal{N}(0, \sigma_u^2 (\mathbf{B}' \mathbf{B})^{-1}) \qquad \mathbf{B} \equiv \mathbf{I}_J - \lambda \mathbf{M} \tag{57}$$

$$y_{ij} = 1 \Longleftrightarrow y_{ij}^* \geq 0 \tag{58}$$

$$y_{ij} = 0 \Longleftrightarrow y_{ij}^* < 0 \tag{59}$$

As discussed in detail above, the DGP may be summarized by the above equations. We vary the parameters of ρ and λ such that ρ, $\lambda \in \{0, 0.3, 0.5\}$. Thus, there are 9 unique combinations of different values for these two parameters. Note that some of these combinations render some special cases of the above DGP. For example, when both ρ and λ are equal to zero the above model becomes a random intercept probit model. When just λ equals zero, our model becomes a SAR probit model with random intercept. We set the values of β_0 and β_1 as -0.5 and 1.0, respectively.[19] We generate values of X_1 from an independent standard normal distribution. The initial values for running the chain are set to 0 for all the parameters. This allows us to assess whether the algorithm allows the iteration to be sampled from values with high posterior densities. We ran 100 trials for each combination of parameters.[20] For reasons of computational demands, we set the number of draws to 1000

[19] β_0 is the intercept. In the appendix, we conduct further simulations where we draw X_1 from a spatially correlated process. $X_1 = (\mathbf{I} - \rho_x \mathbf{W})^{-1} \boldsymbol{\epsilon}$ where $\boldsymbol{\epsilon}$ is white noise.

[20] Some trials were discarded due to numerical instability.

for each trial and discarded the first 200 draws as a burn-in.

Since the conditional distributions of ρ and λ do not have known distributions, we adopt a random walk Metropolis-Hastings for estimating these two parameters. Past research on Bayesian statistics differ on the appropriate acceptance rate for the Metropolis algorithm. We allowed the step size to auto-tune itself for the burn-in period to aim for an acceptance rate of approximately 50% as per the advice of LeSage and Pace (2009). For numerical stability, we logged the acceptance ratio instead of using the ratio directly (Hoff, 2009).

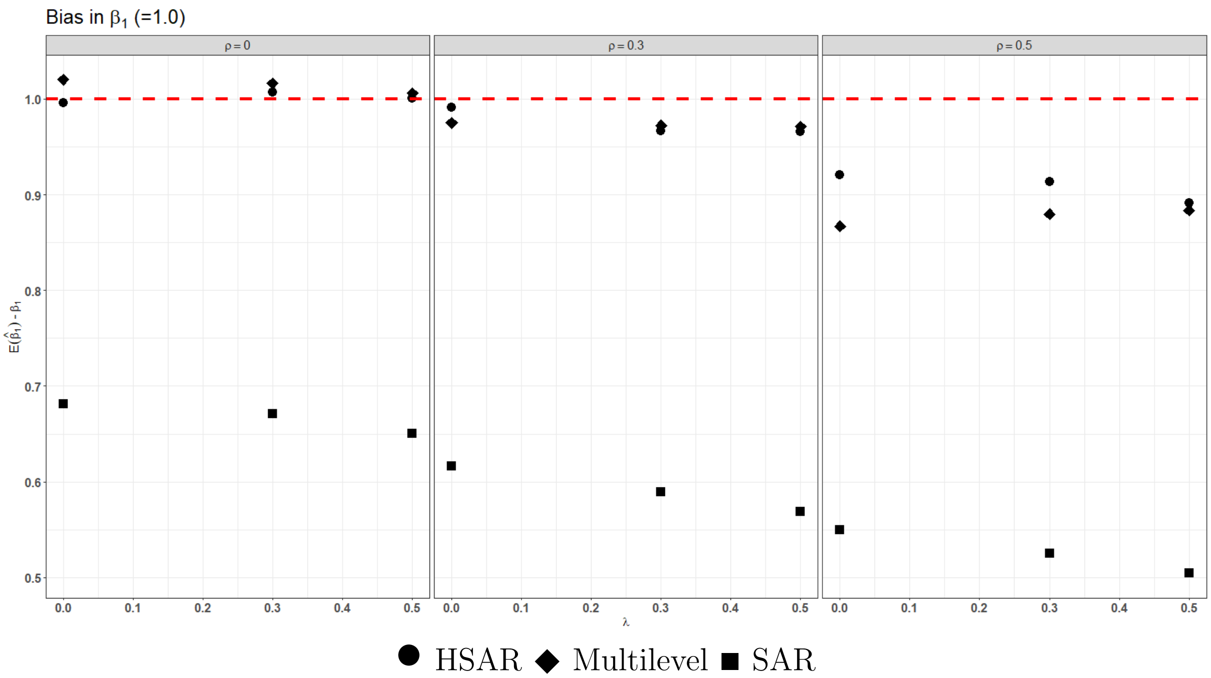
8 Monte Carlo Results

Below, we focus our discussion to the results for the binary probit case. For a given data-generating process, we estimated our proposed HSAR probit model, the traditional SAR probit model and the multilevel random intercept probit model.[21] We focuses our discussion to the results for$\hat{\beta}_1$ and $\hat{\rho}$ below.[22] The dashed red line represents the true value of $\beta_1 = 1.0$. The circle represents the HSAR probit model, the diamond represents the multilevel random intercept probit model,and the traditional SAR probit model. We first present the results for the Monte Carlo simulations with 49 high-level units and 980 low-level units with σ_u^2 set to 1 in Figure **??**.

The results are consistent with our expectations. We see that there is an increasing degree of attenuation bias for the multilevel random intercept probit model as ρ increases in the true DGP because the model is incorrectly assuming that there is no spatial process. In other words, for a given level of λ, there is an increasing degree of attenuation bias for the

[21]We refer the readers to the appendix for the results for the ordered probit model.
[22]More detailed results are presented in the appendix.

Figure 1: Bias in $\hat{\beta}_1$ for $J = 49$, $N = 980$, $\sigma_u^2 = 1.0$



multilevel random intercept probit model as ρ increases. We also see that the SAR probit model induces attenuation bias even when λ equals zero as expected. As discussed above, the traditional SAR probit model does not account for the random intercept and using this model incorrectly to estimate an HSAR process induces an attenuation bias for estimating β_3. We note that even using the HSAR model to estimate an HSAR process induces a noticeable degree of attenuation bias when $\rho = 0.5$. The attenuation bias that we observe here is comparable to those of past studies on spatial probit models (Franzese, Hays and Cook, 2016; Wucherpfennig et al., 2021).

The Monte Carlo results for $\hat{\rho}$ confirm our expectations as shown in Figure **??**. The red dashed line represents the true values of ρ. As discussed above, the random intercept induces attenuation bias for estimating β_3, and this in turn induces an inflationary bias for estimating ρ. In particular, it is worth emphasizing that we observe this phenomenon even when $\lambda = 0$—i.e., there is no spatial process amongst the high-level units.

Figure 2: Bias in $\hat{\rho}$ for $J = 49$, $N = 980$, $\sigma_u^2 = 1.0$

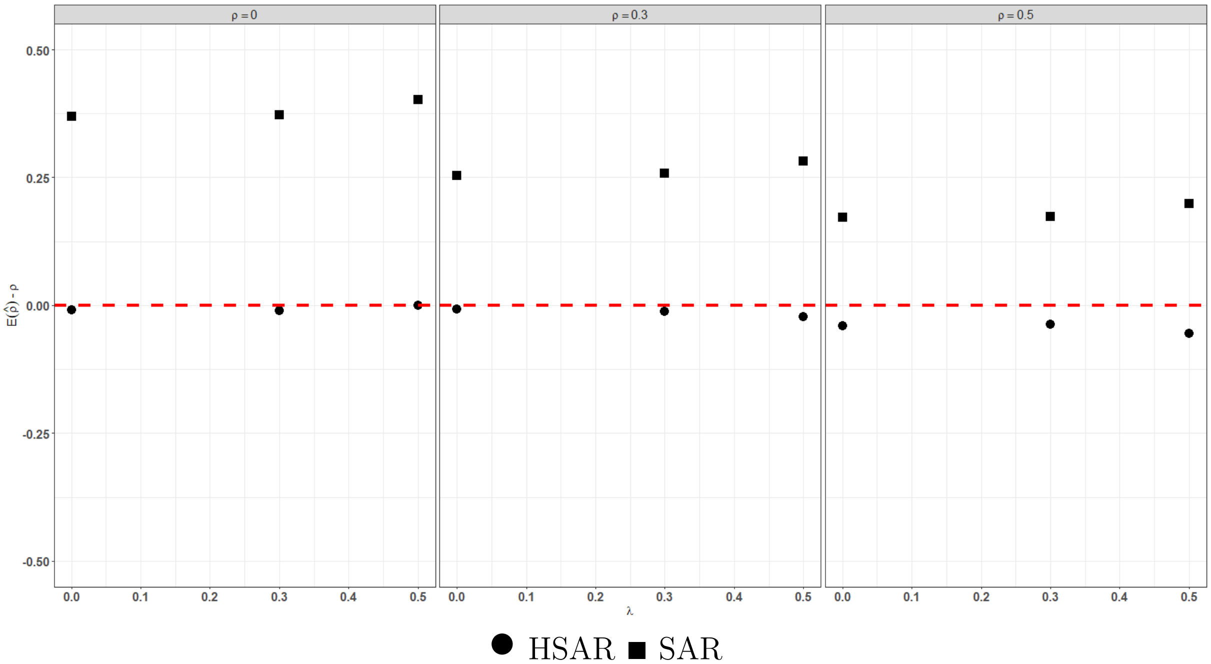
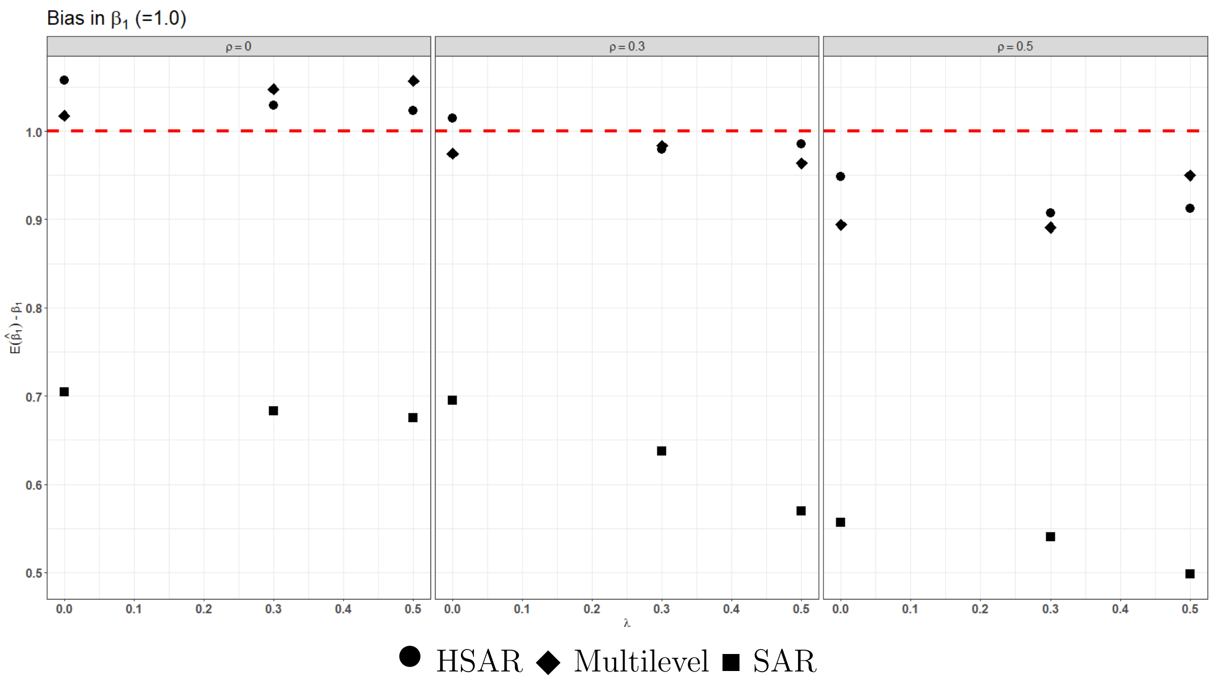


Figure ?? shows the Monte Carlo simulation results when $\sigma_u^2 = 0.5$. We see that the HSAR probit model once again shows the smallest degree of bias for estimating β_3. One difference we observe in Figure ?? is that while the multilevel random intercept probit model performed better than the traditional SAR probit model for all combinations of ρ and λ when $\sigma_u^2 = 1.0$, the multilevel model performs worse than the SAR probit model when $\rho = 0.5$ and $\sigma_u^2 = 0.5$.

[Figure here]

We also present the results when $J = 16$, i.e. when there are 16 high-level units. Figure ?? shows that the multilevel random intercept probit model and the HSAR probit model perform similarly when $\rho = 0$. However, once again, the performance of the multilevel model becomes worse as ρ increases. We also see that, once again, using the traditional SAR probit model induces an attenuation bias in recovering β_3 estimates.

Figure 3: Bias in $\hat{\beta}_1$ for J = 16, N = 320, σ_u^2 = 1.0



Ordered Probit

Extending the algorithm applied above to ordered outcomes is relatively straightforward. In the case of regular spatial ordered probit with three categories, LeSage and Pace (2009) shows that the spatial ordered probit is a straightforward extension of the spatial binary probit model. We can apply the same algorithm as the binary probit case other than minute changes needed for generating $\mathbf{y}^*$ and estimating ϕ_j which represent the cutoff thresholds for categorizing the latent values into different discrete outcomes. Once again, if we conceptualize the outcome as a continuous latent variable, the observation y_{ij} is of category c if

$$y_{ij} = c \quad \text{if} \quad \phi_{c-1} < y_i^* \leq \phi_c \tag{60}$$

LeSage and Pace (2009, 297) notes that for an ordered case of C alternatives, three values of ϕ are fixed, namely $\phi_0 = -\infty$, $\phi_1 = 0$ and $\phi_C = +\infty$ while the thresholds ϕ_c for $c = 2, ..., C - 1$ are to be estimated. The details for estimating these parameters are explained in LeSage and Pace (2009, 297-299). We use the codes from the **spatialprobit** package (Wilhelm and de Matos, 2013) for estimating these cutpoints for our hierarchical ordered spatial probit model.

In the appendix, we conduct additional simulations where $J = 49$ and $N = 980$, $\sigma_u^2 = 1.0$ similar to the binary case. We compare the results from our hierarchical ordered spatial probit model to the ordered spatial probit model implemented with the **spatialprobit** package

(Wilhelm and de Matos, 2013).[23]

Application

We now demonstrate the utility of our model by analyzing the diffusion process of civil rights protests in the United States in the 1960s. There are good theoretical reasons *not* to overlook the (potential) diffusion process of civil rights protests. Theoretically, scholars have debated whether and to what extent protests diffuse in various contexts (e.g., Hale, 2019). In the context of the United States' civil rights protests in the 1960s, sociologists have pointed out various mechanisms through which protests might have spread. For example, Andrews and Biggs (2006) argues that local newspapers played an important role in the diffusion of protests across nearby cities. At the same time, we argue that the potential diffusion process across states has to be taken into account for at least two reasons. First, cities nearby may be dispersed across different states. A cursory glance of the map showing where the sit-ins in the 1960s occurred suggests that there might have been a potential diffusion process for neighboring cities across North Carolina and Virginia. Second, historical accounts suggest that interstate diffusion process might be an important factor to take into account.

We use the dataset provided by Mazumder (2018) and investigate the potential causes of civil rights protests. We take various modeling approaches and compare the coefficient estimates from the different models. The unit of analysis is county in the United States. The dependent variable is whether a civil rights protest took place at least once during the period, coded as 1 if any protest took place and 0 otherwise. For our covariates, we include

[23] We could not find a suitable R package for implementing the ordered probit model.

the percentage of urban population, the percentage of black population, the median age and the median years of school education. We use various model specifications to demonstrate the differences across the models. We first estimate the model using ordinary least squares, then sequentially estimate the model using probit, multilevel random intercept probit, SAR Probit using Bayesian procedure, SAR probit using maximum likelihood procedure, and finally our hierarchical SAR probit. We include state dummies as factor variables for OLS and probit, and as random intercepts for the multilevel random intercept probit model and for our hierarchical SAR probit model.

[Table Here]

9 Conclusion

Our contribution in this paper is our proposed HSAR model which accounts for spatial interdependence at two levels and accounts for categorical outcomes. Our results from a series of Monte Carlo experiments demonstrate the viability of our HSAR model as general model that can account for spatial processes that have interdependence at both one or two levels.

Our proposed method and Monte Carlo experiments have certain limitations that need addressing. The first is assessing the performance of the HSAR model while varying both the number of lower- and higher-level units. In the applied example with a continuous outcome variable, Dong and Harris (2015) has approximately 1100 lower level units distributed over 100 high level units. Intuitively, there has to be a sufficient number of units at both levels. For example, adopting this approach crudely with states grouped into continents would probably not render better inferences. As with other applications in statistics, having a low number of units and insufficient variation in the outcome is likely to lead to erroneous inferences. We believe our approach would be the most useful in comparative politics settings where the researcher is interested in modeling diffusion processes at two levels within a country.

References

Albert, James H and Siddhartha Chib. 1993. "Bayesian Analysis of Binary and Polychotomous Response Data." *Journal of the American Statistical Association* 88(422):669–679.

Andrews, Kenneth T and Michael Biggs. 2006. "The Dynamics of Protest Diffusion: Movement Organizations, Social Networks, and News Media in the 1960 Sit-ins." *American Sociological Review* 71(5):752–777.

Angrist, Joshua, Eric Bettinger and Michael Kremer. 2006. "Long-Term Educational Consequences of Secondary School Vouchers: Evidence from Administrative Records in Colombia." *American Economic Review* 96(3):847–862.

Beron, Kurt J and Wim PM Vijverberg. 2004. Probit in a Spatial Context: A Monte Carlo Analysis. In *Advances in Spatial Econometrics: Methodology, Tools and Applications*. Springer pp. 169–195.

Carpenter, Bob, Andrew Gelman, Matthew D Hoffman, Daniel Lee, Ben Goodrich, Michael Betancourt, Marcus A Brubaker, Jiqiang Guo, Peter Li and Allen Riddell. 2017. "Stan: A Probabilistic Programming Language." *Journal of statistical software* 76.

Cramer, Jan Salomon. 2003. *Logit mMdels from Economics and Other Fields*. Cambridge University Press.

Dong, Guanpeng and Richard Harris. 2015. "Spatial Autoregressive Models for Geographically Hierarchical Data Structures." *Geographical Analysis* 47(2):173–191.

Franzese, Robert J, Jude C Hays and Scott J Cook. 2016. "Spatial and Spatiotemporal-autoregressive Probit Models of Interdependent Binary Outcomes." *Political Science Research and Methods* 4(1):151–173.

Gelman, Andrew and Jennifer Hill. 2006. *Data Analysis using Regression and Multilevel/Hierarchical models*. Cambridge university press.

Gelman, Andrew, John B Carlin, Hal S Stern, David B Dunson, Aki Vehtari and Donald B Rubin. 2013. *Bayesian Data Analysis*. CRC press.

George, Justin and Todd Sandler. 2018. "Demand for Military Spending in NATO, 1968–2015: A Spatial Panel Approach." *European Journal of Political Economy* 53:222–236.

Geweke, John. 1991. Efficient Simulation from the Multivariate Normal and Student-t Distributions subject to Linear Constraints and the Evaluation of Constraint Probabilities. In *Computing Science and Statistics: Proceedings of the 23rd Symposium on the Interface*. Vol. 571 Fairfax, Virginia: Interface Foundation of North America, Inc p. 578.

Ghitza, Yair and Andrew Gelman. 2013. "Deep Interactions with MRP: Election Turnout and Voting Patterns among Small Electoral Subgroups." *American Journal of Political Science* 57(3):762–776.

Hale, Henry E. 2019. "How should We Now Conceptualize Protest, Diffusion, and Regime Change?" *Journal of Conflict Resolution* 63(10):2402–2415.

Halleck Vega, Solmaria and J Paul Elhorst. 2015. "The SLX Model." *Journal of Regional Science* 55(3):339–363.

Hoff, Peter D. 2009. *A First Course in Bayesian Statistical Methods*. Springer.

Jackman, Simon. 2009. *Bayesian Analysis for the Social Sciences*. John Wiley & Sons.

Kotecha, Jayesh H and Petar M Djuric. 1999. Gibbs Sampling Approach for Generation of Truncated Multivariate Gaussian Random Variables. In *1999 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings. ICASSP99 (Cat. No. 99CH36258)*. Vol. 3 IEEE pp. 1757–1760.

LeSage, James and Robert Kelley Pace. 2009. *Introduction to Spatial Econometrics*. Chapman and Hall/CRC.

Lynch, Scott M. 2007. *Introduction to Applied Bayesian Statistics and Estimation for Social Scientists*. Vol. 1 Springer.

Mazumder, Soumyajit. 2018. "The Persistent Effect of US Civil Rights Protests on Political Attitudes." *American Journal of Political Science* 62(4):922–935.

Neuhaus, John M and Nicholas P Jewell. 1993. "A Geometric Approach to Assess Bias due to Omitted Covariates in Generalized Linear Models." *Biometrika* 80(4):807–815.

Simmons, Beth A and Zachary Elkins. 2004. "The Globalization of Liberalization: Policy Diffusion in the International Political Economy." *American Political Science Review* 98(1):171–189.

Smith, Tony E and James P LeSage. 2004. A Bayesian Probit Model with Spatial Dependencies. In *Spatial and Spatiotemporal Econometrics*. Emerald Group Publishing Limited pp. 127–160.

Tanner, Martin A and Wing Hung Wong. 1987. “The Calculation of Posterior Distributions by Data Augmentation.” *Journal of the American Statistical Association* 82(398):528–540.

Wilhelm, Stefan and BG Manjunath. 2010. “tmvtnorm: A package for the truncated multivariate normal distribution.” *sigma* 2(2):1–25.

Wilhelm, Stefan and Miguel Godinho de Matos. 2013. “Estimating Spatial Probit Models in R.” *R Journal* 5(1):130.

Williams, Laron K and Guy D Whitten. 2015. “Don’t Stand So Close to Me: Spatial Contagion Effects and Party Competition.” *American Journal of Political Science* 59(2):309–325.

Wolf, Levi John, Luc Anselin and Daniel Arribas-Bel. 2018. “Stochastic Efficiency of Bayesian Markov chain Monte Carlo in Spatial Econometric Models: An Empirical Comparison of Exact Sampling Methods.” *Geographical Analysis* 50(1):97–119.

Wucherpfennig, Julian, Aya Kachi, Nils-Christian Bormann and Philipp Hunziker. 2021. “A Fast Estimator for Binary Choice Models with Spatial, Temporal, and Spatio-Temporal Interdependence.” *Political Analysis* 29(4):570–576.

Online Appendix for

Modeling Hierarchical Spatial Interdependence for Limited Dependent Variables

Binary Probit $J = 49$, $N = 980$, $\sigma_u^2 = 1.0$, $\rho_x = 0.0$

Experiment 1	HSAR				SAR			Multilevel	
$\rho = 0.0$, $\lambda = 0.0$	ρ	λ	β_0	β_1	ρ	β_0	β_1	β_0	β_1
Bias	-0.009	-0.03	-0.005	-0.004	0.369	0.268	-0.319	-0.007	0.02
SD	0.068	0.214	0.173	0.076	0.054	0.076	0.078	0.147	0.086
RMSE	0.069	0.216	0.173	0.076	0.373	0.278	0.328	0.147	0.089

Experiment 2	HSAR				SAR			Multilevel	
$\rho = 0.0$, $\lambda = 0.3$	ρ	λ	β_0	β_1	ρ	β_0	β_1	β_0	β_1
Bias	-0.011	-0.044	0.006	0.007	0.372	0.262	-0.329	-0.014	0.016
SD	0.075	0.199	0.294	0.079	0.069	0.115	0.074	0.216	0.074
RMSE	0.076	0.204	0.294	0.079	0.378	0.287	0.337	0.216	0.076

Experiment 3	HSAR				SAR			Multilevel	
$\rho = 0.0$, $\lambda = 0.5$	ρ	λ	β_0	β_1	ρ	β_0	β_1	β_0	β_1
Bias	0	-0.043	-0.037	0.001	0.402	0.275	-0.35	-0.013	0.006
SD	0.072	0.143	0.338	0.076	0.055	0.121	0.081	0.263	0.068
RMSE	0.072	0.149	0.34	0.076	0.405	0.3	0.359	0.264	0.069

Experiment 4	HSAR				SAR			Multilevel	
$\rho = 0.3$, $\lambda = 0.0$	ρ	λ	β_0	β_1	ρ	β_0	β_1	β_0	β_1
Bias	-0.007	-0.027	0.028	-0.009	0.254	0.283	-0.384	-0.194	-0.025
SD	0.061	0.224	0.149	0.072	0.039	0.075	0.082	0.22	0.075
RMSE	0.061	0.225	0.152	0.073	0.257	0.292	0.392	0.293	0.079

Experiment 5	HSAR				SAR			Multilevel	
ρ = 0.3, λ = 0.3	ρ	λ	β_0	β_1	ρ	β_0	β_1	β_0	β_1
Bias	-0.012	-0.057	-0.008	-0.033	0.259	0.272	-0.41	-0.217	-0.028
SD	0.055	0.19	0.222	0.08	0.039	0.09	0.076	0.27	0.084
RMSE	0.057	0.199	0.222	0.087	0.261	0.286	0.417	0.347	0.088

Experiment 6	HSAR				SAR			Multilevel	
ρ = 0.3, λ = 0.5	ρ	λ	β_0	β_1	ρ	β_0	β_1	β_0	β_1
Bias	-0.023	-0.036	0.082	-0.034	0.282	0.279	-0.431	-0.256	-0.029
SD	0.057	0.158	0.331	0.071	0.049	0.136	0.087	0.408	0.081
RMSE	0.061	0.162	0.341	0.078	0.286	0.31	0.44	0.482	0.086

Experiment 7	HSAR				SAR			Multilevel	
ρ = 0.5, λ = 0.0	ρ	λ	β_0	β_1	ρ	β_0	β_1	β_0	β_1
Bias	-0.04	-0.023	0.036	-0.079	0.173	0.283	-0.45	-0.303	-0.133
SD	0.05	0.203	0.156	0.084	0.036	0.072	0.102	0.245	0.081
RMSE	0.064	0.205	0.16	0.116	0.176	0.292	0.461	0.39	0.156

Experiment 8	HSAR				SAR			Multilevel	
ρ = 0.5, λ = 0.3	ρ	λ	β_0	β_1	ρ	β_0	β_1	β_0	β_1
Bias	-0.037	-0.08	0.042	-0.086	0.174	0.276	-0.475	-0.322	-0.12
SD	0.047	0.197	0.209	0.088	0.038	0.094	0.076	0.359	0.073
RMSE	0.059	0.212	0.213	0.123	0.178	0.292	0.481	0.482	0.141

Experiment 9	HSAR				SAR			Multilevel	
ρ = 0.5, λ = 0.5	ρ	λ	β_0	β_1	ρ	β_0	β_1	β_0	β_1
Bias	-0.055	-0.098	0.034	-0.109	0.199	0.29	-0.495	-0.403	-0.116
SD	0.049	0.167	0.276	0.094	0.042	0.129	0.104	0.516	0.092
RMSE	0.073	0.194	0.278	0.144	0.203	0.318	0.506	0.655	0.149

Binary Probit $J = 49$, $N = 980$, $\sigma_u^2 = 1.0$, $\rho_x = 0.3$

Experiment 1	HSAR				SAR			Multilevel	
$\rho = 0.0$, $\lambda = 0.0$	ρ	λ	β_0	β_1	ρ	β_0	β_1	β_0	β_1
Bias	-0.003	-0.048	-0.015	0.011	0.348	0.258	-0.383	0.018	0.016
SD	0.07	0.229	0.161	0.077	0.056	0.085	0.075	0.164	0.076
RMSE	0.07	0.234	0.162	0.078	0.353	0.272	0.39	0.165	0.078

Experiment 2	HSAR				SAR			Multilevel	
$\rho = 0.0$, $\lambda = 0.3$	ρ	λ	β_0	β_1	ρ	β_0	β_1	β_0	β_1
Bias	-0.016	-0.083	-0.029	-0.001	0.371	0.239	-0.372	-0.04	0.023
SD	0.068	0.19	0.25	0.076	0.059	0.114	0.077	0.235	0.074
RMSE	0.07	0.207	0.252	0.076	0.376	0.265	0.38	0.238	0.078

Experiment 3	HSAR				SAR			Multilevel	
$\rho = 0.0$, $\lambda = 0.5$	ρ	λ	β_0	β_1	ρ	β_0	β_1	β_0	β_1
Bias	-0.011	-0.034	-0.009	0.004	0.403	0.268	-0.426	-0.033	0.013
SD	0.079	0.158	0.434	0.073	0.05	0.137	0.077	0.31	0.073
RMSE	0.079	0.162	0.434	0.073	0.406	0.301	0.433	0.312	0.074

Experiment 4	HSAR				SAR			Multilevel	
$\rho = 0.3$, $\lambda = 0.0$	ρ	λ	β_0	β_1	ρ	β_0	β_1	β_0	β_1
Bias	-0.01	-0.055	-0.002	-0.016	0.25	0.239	-0.434	-0.168	0.042
SD	0.062	0.218	0.155	0.071	0.038	0.066	0.065	0.173	0.091
RMSE	0.063	0.225	0.155	0.072	0.253	0.248	0.439	0.241	0.1

Experiment 5	HSAR				SAR			Multilevel	
$\rho = 0.3,\ \lambda = 0.3$	ρ	λ	β_0	β_1	ρ	β_0	β_1	β_0	β_1
Bias	-0.006	-0.07	0.027	-0.014	0.25	0.269	-0.448	-0.2	0.02
SD	0.049	0.2	0.203	0.077	0.044	0.107	0.083	0.306	0.083
RMSE	0.049	0.212	0.205	0.078	0.254	0.29	0.455	0.366	0.085

Experiment 6	HSAR				SAR			Multilevel	
$\rho = 0.3,\ \lambda = 0.5$	ρ	λ	β_0	β_1	ρ	β_0	β_1	β_0	β_1
Bias	-0.017	-0.076	0.017	-0.022	0.286	0.291	-0.498	-0.169	0.038
SD	0.056	0.194	0.351	0.086	0.044	0.132	0.08	0.423	0.075
RMSE	0.058	0.208	0.352	0.089	0.289	0.32	0.505	0.456	0.084

Experiment 7	HSAR				SAR			Multilevel	
$\rho = 0.5,\ \lambda = 0.0$	ρ	λ	β_0	β_1	ρ	β_0	β_1	β_0	β_1
Bias	-0.039	0.01	0.073	-0.095	0.15	0.249	-0.478	-0.37	-0.027
SD	0.044	0.214	0.138	0.068	0.037	0.072	0.088	0.242	0.086
RMSE	0.059	0.214	0.156	0.116	0.154	0.259	0.486	0.443	0.09

Experiment 8	HSAR				SAR			Multilevel	
$\rho = 0.5,\ \lambda = 0.3$	ρ	λ	β_0	β_1	ρ	β_0	β_1	β_0	β_1
Bias	-0.043	-0.06	0.044	-0.095	0.18	0.282	-0.517	-0.336	-0.036
SD	0.044	0.221	0.222	0.07	0.04	0.1	0.093	0.359	0.083
RMSE	0.061	0.229	0.226	0.118	0.185	0.299	0.526	0.491	0.09

Experiment 9	HSAR				SAR			Multilevel	
$\rho = 0.5,\ \lambda = 0.5$	ρ	λ	β_0	β_1	ρ	β_0	β_1	β_0	β_1
Bias	-0.045	-0.073	0.034	-0.104	0.19	0.289	-0.558	-0.357	-0.043
SD	0.048	0.164	0.335	0.083	0.04	0.118	0.081	0.47	0.085
RMSE	0.066	0.179	0.337	0.133	0.195	0.312	0.564	0.59	0.095

Binary Probit $J = 16$, $N = 320$, $\sigma_u^2 = 1.0$, $\rho_x = 0.0$

Experiment 1	HSAR				SAR			Multilevel	
$\rho = 0.0$, $\lambda = 0.0$	ρ	λ	β_0	β_1	ρ	β_0	β_1	β_0	β_1
Bias	-0.027	-0.059	-0.024	0.058	0.345	0.208	-0.296	-0.051	0.018
SD	0.126	0.285	0.324	0.14	0.109	0.146	0.144	0.242	0.136
RMSE	0.129	0.291	0.325	0.151	0.362	0.255	0.329	0.248	0.137

Experiment 2	HSAR				SAR			Multilevel	
$\rho = 0.0$, $\lambda = 0.3$	ρ	λ	β_0	β_1	ρ	β_0	β_1	β_0	β_1
Bias	-0.03	-0.159	0.004	0.03	0.365	0.245	-0.317	-0.027	0.048
SD	0.125	0.313	0.509	0.136	0.111	0.193	0.131	0.349	0.144
RMSE	0.128	0.351	0.509	0.139	0.381	0.312	0.343	0.35	0.152

Experiment 3	HSAR				SAR			Multilevel	
$\rho = 0.0$, $\lambda = 0.5$	ρ	λ	β_0	β_1	ρ	β_0	β_1	β_0	β_1
Bias	-0.012	-0.161	0.007	0.023	0.384	0.277	-0.324	0.035	0.057
SD	0.12	0.264	0.732	0.137	0.132	0.276	0.137	0.546	0.122
RMSE	0.121	0.309	0.732	0.139	0.406	0.391	0.352	0.547	0.135

Experiment 4	HSAR				SAR			Multilevel	
$\rho = 0.3$, $\lambda = 0.0$	ρ	λ	β_0	β_1	ρ	β_0	β_1	β_0	β_1
Bias	-0.044	-0.02	-0.09	0.015	0.25	0.245	-0.305	-0.075	-0.026
SD	0.088	0.3	0.353	0.12	0.068	0.133	0.149	0.383	0.123
RMSE	0.098	0.3	0.364	0.121	0.259	0.279	0.34	0.39	0.126

Experiment 5	HSAR				SAR			Multilevel	
$\rho = 0.3,\ \lambda = 0.3$	ρ	λ	β_0	β_1	ρ	β_0	β_1	β_0	β_1
Bias	-0.031	-0.146	0.022	-0.02	0.265	0.281	-0.363	-0.157	-0.016
SD	0.097	0.308	0.446	0.13	0.08	0.192	0.15	0.538	0.139
RMSE	0.102	0.341	0.446	0.132	0.276	0.341	0.393	0.561	0.14

Experiment 6	HSAR				SAR			Multilevel	
$\rho = 0.3,\ \lambda = 0.5$	ρ	λ	β_0	β_1	ρ	β_0	β_1	β_0	β_1
Bias	-0.044	-0.178	-0.018	-0.015	0.253	0.282	-0.43	-0.135	-0.036
SD	0.122	0.305	0.671	0.157	0.078	0.226	0.174	0.822	0.141
RMSE	0.13	0.353	0.671	0.158	0.264	0.361	0.464	0.833	0.146

Experiment 7	HSAR				SAR			Multilevel	
$\rho = 0.5,\ \lambda = 0.0$	ρ	λ	β_0	β_1	ρ	β_0	β_1	β_0	β_1
Bias	-0.051	-0.003	0.001	-0.051	0.149	0.266	-0.444	-0.437	-0.106
SD	0.081	0.306	0.341	0.125	0.074	0.127	0.135	0.493	0.143
RMSE	0.096	0.306	0.341	0.135	0.167	0.294	0.464	0.659	0.178

Experiment 8	HSAR				SAR			Multilevel	
$\rho = 0.5,\ \lambda = 0.3$	ρ	λ	β_0	β_1	ρ	β_0	β_1	β_0	β_1
Bias	-0.056	-0.134	-0.001	-0.093	0.162	0.271	-0.46	-0.343	-0.109
SD	0.086	0.315	0.465	0.141	0.075	0.181	0.15	0.7	0.137
RMSE	0.103	0.342	0.465	0.169	0.179	0.326	0.483	0.78	0.175

Experiment 9	HSAR				SAR			Multilevel	
$\rho = 0.5,\ \lambda = 0.5$	ρ	λ	β_0	β_1	ρ	β_0	β_1	β_0	β_1
Bias	-0.085	-0.172	0.003	-0.087	0.191	0.299	-0.502	-0.234	-0.05
SD	0.109	0.273	0.71	0.155	0.074	0.236	0.177	1.337	0.156
RMSE	0.139	0.323	0.71	0.178	0.205	0.381	0.532	1.357	0.164

Binary Probit $J = 16$, $N = 320$, $\sigma_u^2 = 0.5$, $\rho_x = 0.0$

Experiment 1	HSAR				SAR			Multilevel	
$\rho = 0.0,\ \lambda = 0.0$	ρ	λ	β_0	β_1	ρ	β_0	β_1	β_0	β_1
Bias	-0.024	-0.001	0.003	0.021	0.227	0.135	-0.174	-0.038	0.021
SD	0.12	0.298	0.285	0.116	0.122	0.141	0.133	0.177	0.131
RMSE	0.123	0.298	0.285	0.118	0.258	0.195	0.219	0.181	0.132

Experiment 2	HSAR				SAR			Multilevel	
$\rho = 0.0,\ \lambda = 0.3$	ρ	λ	β_0	β_1	ρ	β_0	β_1	β_0	β_1
Bias	-0.019	-0.134	0.012	0.031	0.246	0.169	-0.194	-0.022	0.039
SD	0.124	0.299	0.35	0.133	0.125	0.183	0.126	0.25	0.136
RMSE	0.125	0.328	0.35	0.137	0.276	0.25	0.232	0.251	0.141

Experiment 3	HSAR				SAR			Multilevel	
$\rho = 0.0,\ \lambda = 0.5$	ρ	λ	β_0	β_1	ρ	β_0	β_1	β_0	β_1
Bias	-0.03	-0.228	-0.01	0.044	0.269	0.2	-0.197	0.02	0.037
SD	0.125	0.303	0.539	0.136	0.144	0.258	0.12	0.383	0.12
RMSE	0.129	0.379	0.539	0.143	0.305	0.327	0.231	0.384	0.126

Experiment 4	HSAR				SAR			Multilevel	
$\rho = 0.3,\ \lambda = 0.0$	ρ	λ	β_0	β_1	ρ	β_0	β_1	β_0	β_1
Bias	-0.025	-0.063	0.01	-0.005	0.178	0.171	-0.182	-0.091	-0.033
SD	0.102	0.321	0.425	0.135	0.068	0.132	0.147	0.275	0.126
RMSE	0.105	0.327	0.425	0.135	0.191	0.216	0.234	0.29	0.13

Experiment 5	HSAR				SAR			Multilevel	
ρ = 0.3, λ = 0.3	ρ	λ	β_0	β_1	ρ	β_0	β_1	β_0	β_1
Bias	-0.025	-0.143	-0.008	0	0.196	0.213	-0.237	-0.164	-0.01
SD	0.112	0.281	0.32	0.13	0.078	0.175	0.142	0.377	0.133
RMSE	0.115	0.315	0.32	0.13	0.211	0.275	0.276	0.411	0.133

Experiment 6	HSAR				SAR			Multilevel	
ρ = 0.3, λ = 0.5	ρ	λ	β_0	β_1	ρ	β_0	β_1	β_0	β_1
Bias	-0.026	-0.168	-0.007	0.012	0.179	0.207	-0.283	-0.125	-0.027
SD	0.094	0.275	0.416	0.144	0.075	0.215	0.168	0.511	0.138
RMSE	0.097	0.322	0.416	0.145	0.194	0.298	0.329	0.526	0.14

Experiment 7	HSAR				SAR			Multilevel	
ρ = 0.5, λ = 0.0	ρ	λ	β_0	β_1	ρ	β_0	β_1	β_0	β_1
Bias	-0.043	-0.002	-0.034	-0.058	0.093	0.188	-0.304	-0.431	-0.102
SD	0.081	0.306	0.241	0.139	0.073	0.122	0.124	0.337	0.116
RMSE	0.092	0.306	0.244	0.151	0.118	0.224	0.328	0.547	0.154

Experiment 8	HSAR				SAR			Multilevel	
ρ = 0.5, λ = 0.3	ρ	λ	β_0	β_1	ρ	β_0	β_1	β_0	β_1
Bias	-0.038	-0.173	0.023	-0.02	0.103	0.198	-0.323	-0.351	-0.067
SD	0.069	0.261	0.324	0.141	0.072	0.174	0.148	0.427	0.141
RMSE	0.079	0.313	0.324	0.143	0.126	0.264	0.356	0.553	0.156

Experiment 9	HSAR				SAR			Multilevel	
ρ = 0.5, λ = 0.5	ρ	λ	β_0	β_1	ρ	β_0	β_1	β_0	β_1
Bias	-0.066	-0.236	-0.076	-0.074	0.124	0.225	-0.361	-0.215	-0.031
SD	0.097	0.287	0.495	0.143	0.069	0.226	0.17	0.703	0.134
RMSE	0.118	0.371	0.501	0.161	0.142	0.319	0.399	0.736	0.138

Binary Probit $J = 16$, $N = 320$, $\sigma_u^2 = 1.0$, $\rho_x = 0.3$

Experiment 1	HSAR				SAR			Multilevel	
$\rho = 0.0$, $\lambda = 0.0$	ρ	λ	β_0	β_1	ρ	β_0	β_1	β_0	β_1
Bias	-0.03	-0.01	-0.047	0.03	0.324	0.269	-0.34	0.025	0.011
SD	0.115	0.27	0.327	0.124	0.11	0.168	0.131	0.29	0.123
RMSE	0.118	0.27	0.33	0.128	0.342	0.317	0.365	0.291	0.123

Experiment 2	HSAR				SAR			Multilevel	
$\rho = 0.0$, $\lambda = 0.3$	ρ	λ	β_0	β_1	ρ	β_0	β_1	β_0	β_1
Bias	-0.029	-0.148	-0.084	0.037	0.344	0.245	-0.37	-0.004	0.026
SD	0.134	0.3	0.521	0.126	0.098	0.199	0.136	0.38	0.129
RMSE	0.137	0.335	0.528	0.132	0.358	0.316	0.394	0.38	0.131

Experiment 3	HSAR				SAR			Multilevel	
$\rho = 0.0$, $\lambda = 0.5$	ρ	λ	β_0	β_1	ρ	β_0	β_1	β_0	β_1
Bias	-0.034	-0.2	0.02	0.033	0.33	0.254	-0.355	-0.013	0.03
SD	0.114	0.259	0.64	0.153	0.117	0.251	0.137	0.478	0.141
RMSE	0.119	0.327	0.64	0.157	0.35	0.357	0.38	0.479	0.144

Experiment 4	HSAR				SAR			Multilevel	
$\rho = 0.3$, $\lambda = 0.0$	ρ	λ	β_0	β_1	ρ	β_0	β_1	β_0	β_1
Bias	-0.036	-0.003	-0.038	-0.004	0.21	0.238	-0.409	-0.277	0.059
SD	0.104	0.298	0.503	0.144	0.085	0.168	0.145	0.415	0.168
RMSE	0.11	0.298	0.504	0.144	0.227	0.291	0.434	0.499	0.178

Experiment 5	HSAR				SAR			Multilevel	
$\rho = 0.3,\ \lambda = 0.3$	ρ	λ	β_0	β_1	ρ	β_0	β_1	β_0	β_1
Bias	-0.049	-0.148	-0.018	-0.002	0.234	0.264	-0.418	-0.178	0.028
SD	0.115	0.313	0.624	0.134	0.089	0.211	0.129	0.477	0.144
RMSE	0.125	0.346	0.624	0.134	0.25	0.338	0.438	0.509	0.147

Experiment 6	HSAR				SAR			Multilevel	
$\rho = 0.3,\ \lambda = 0.5$	ρ	λ	β_0	β_1	ρ	β_0	β_1	β_0	β_1
Bias	-0.029	-0.127	0.049	-0.009	0.254	0.258	-0.442	-0.22	0.029
SD	0.106	0.25	0.721	0.144	0.077	0.263	0.15	0.966	0.163
RMSE	0.11	0.28	0.723	0.144	0.265	0.368	0.466	0.991	0.166

Experiment 7	HSAR				SAR			Multilevel	
$\rho = 0.5,\ \lambda = 0.0$	ρ	λ	β_0	β_1	ρ	β_0	β_1	β_0	β_1
Bias	-0.062	0.024	0.01	-0.065	0.164	0.251	-0.471	-0.454	-0.006
SD	0.072	0.286	0.331	0.143	0.077	0.162	0.132	0.511	0.165
RMSE	0.095	0.287	0.332	0.157	0.182	0.299	0.489	0.684	0.165

Experiment 8	HSAR				SAR			Multilevel	
$\rho = 0.5,\ \lambda = 0.3$	ρ	λ	β_0	β_1	ρ	β_0	β_1	β_0	β_1
Bias	-0.059	-0.124	0.006	-0.091	0.16	0.275	-0.475	-0.247	-0.036
SD	0.084	0.293	0.438	0.127	0.067	0.178	0.208	0.677	0.152
RMSE	0.103	0.318	0.438	0.156	0.174	0.327	0.519	0.721	0.156

Experiment 9	HSAR				SAR			Multilevel	
$\rho = 0.5,\ \lambda = 0.5$	ρ	λ	β_0	β_1	ρ	β_0	β_1	β_0	β_1
Bias	-0.058	-0.223	0.058	-0.082	0.172	0.244	-0.488	-0.485	0.01
SD	0.071	0.282	0.497	0.127	0.081	0.256	0.201	1.165	0.147
RMSE	0.091	0.36	0.5	0.152	0.191	0.354	0.528	1.261	0.148

Binary Probit $J = 16$, $N = 320$, $\sigma_u^2 = 0.5$, $\rho_x = 0.3$

Experiment 1	HSAR				SAR			Multilevel	
$\rho = 0.0$, $\lambda = 0.0$	ρ	λ	β_0	β_1	ρ	β_0	β_1	β_0	β_1
Bias	-0.028	0.002	-0.004	0.026	0.215	0.188	-0.205	0.015	0.018
SD	0.113	0.288	0.237	0.126	0.12	0.166	0.116	0.217	0.111
RMSE	0.116	0.288	0.237	0.129	0.246	0.251	0.235	0.217	0.113

Experiment 2	HSAR				SAR			Multilevel	
$\rho = 0.0$, $\lambda = 0.3$	ρ	λ	β_0	β_1	ρ	β_0	β_1	β_0	β_1
Bias	-0.011	-0.116	-0.061	0.023	0.22	0.169	-0.237	0.001	0.012
SD	0.106	0.287	0.422	0.128	0.115	0.19	0.129	0.283	0.124
RMSE	0.107	0.309	0.426	0.13	0.248	0.254	0.27	0.283	0.124

Experiment 3	HSAR				SAR			Multilevel	
$\rho = 0.0$, $\lambda = 0.5$	ρ	λ	β_0	β_1	ρ	β_0	β_1	β_0	β_1
Bias	0.014	-0.176	0.014	0.026	0.21	0.165	-0.203	-0.021	0.045
SD	0.097	0.252	0.47	0.134	0.125	0.242	0.146	0.346	0.138
RMSE	0.098	0.307	0.471	0.136	0.245	0.293	0.25	0.346	0.145

Experiment 4	HSAR				SAR			Multilevel	
$\rho = 0.3$, $\lambda = 0.0$	ρ	λ	β_0	β_1	ρ	β_0	β_1	β_0	β_1
Bias	-0.022	-0.023	-0.02	0.011	0.132	0.158	-0.267	-0.268	0.062
SD	0.109	0.263	0.26	0.129	0.077	0.163	0.14	0.303	0.132
RMSE	0.111	0.264	0.261	0.129	0.153	0.227	0.301	0.404	0.145

Experiment 5	HSAR				SAR			Multilevel	
$\rho = 0.3,\ \lambda = 0.3$	ρ	λ	β_0	β_1	ρ	β_0	β_1	β_0	β_1
Bias	-0.03	-0.177	-0.043	0.032	0.156	0.187	-0.277	-0.195	0.034
SD	0.092	0.308	0.36	0.139	0.091	0.2	0.128	0.352	0.137
RMSE	0.096	0.355	0.362	0.142	0.18	0.274	0.306	0.402	0.141

Experiment 6	HSAR				SAR			Multilevel	
$\rho = 0.3,\ \lambda = 0.5$	ρ	λ	β_0	β_1	ρ	β_0	β_1	β_0	β_1
Bias	-0.016	-0.212	-0.057	0.013	0.177	0.181	-0.296	-0.182	0.012
SD	0.101	0.279	0.516	0.145	0.089	0.263	0.14	0.52	0.12
RMSE	0.102	0.35	0.519	0.146	0.198	0.319	0.328	0.551	0.12

Experiment 7	HSAR				SAR			Multilevel	
$\rho = 0.5,\ \lambda = 0.0$	ρ	λ	β_0	β_1	ρ	β_0	β_1	β_0	β_1
Bias	-0.059	-0.045	-0.046	-0.025	0.096	0.176	-0.332	-0.452	0.021
SD	0.086	0.302	0.266	0.137	0.065	0.156	0.132	0.367	0.154
RMSE	0.105	0.305	0.27	0.139	0.116	0.235	0.357	0.582	0.156

Experiment 8	HSAR				SAR			Multilevel	
$\rho = 0.5,\ \lambda = 0.3$	ρ	λ	β_0	β_1	ρ	β_0	β_1	β_0	β_1
Bias	-0.045	-0.15	-0.014	-0.037	0.091	0.196	-0.339	-0.252	-0.025
SD	0.068	0.283	0.314	0.133	0.068	0.171	0.197	0.452	0.145
RMSE	0.082	0.32	0.314	0.138	0.114	0.26	0.392	0.518	0.147

Experiment 9	HSAR				SAR			Multilevel	
$\rho = 0.5,\ \lambda = 0.5$	ρ	λ	β_0	β_1	ρ	β_0	β_1	β_0	β_1
Bias	-0.063	-0.208	-0.045	-0.052	0.103	0.17	-0.348	-0.418	0.016
SD	0.095	0.313	0.492	0.144	0.071	0.241	0.192	0.808	0.123
RMSE	0.114	0.376	0.494	0.153	0.125	0.295	0.398	0.909	0.124

Ordered Probit $J = 49$, $N = 980$, $\sigma_u^2 = 1.0$, $\rho_x = 0.0$

Experiment 1	HSAR			SAR		Multilevel
$\rho = 0.0$, $\lambda = 0.0$	ρ	λ	β_1	ρ	β_1	β_1
Bias	-0.001	-0.038	0.009	0.367	-0.339	0.001
SD	0.058	0.222	0.075	0.065	0.047	0.071
RMSE	0.058	0.226	0.076	0.372	0.343	0.071

Experiment 2	HSAR			SAR		Multilevel
$\rho = 0.0$, $\lambda = 0.3$	ρ	λ	β_1	ρ	β_1	β_1
Bias	-0.017	-0.065	-0.014	0.37	-0.35	0
SD	0.063	0.208	0.075	0.072	0.044	0.062
RMSE	0.065	0.218	0.077	0.377	0.353	0.062

Experiment 3	HSAR			SAR		Multilevel
$\rho = 0.0$, $\lambda = 0.5$	ρ	λ	β_1	ρ	β_1	β_1
Bias	-0.005	-0.096	-0.013	0.423	-0.375	-0.006
SD	0.055	0.168	0.072	0.076	0.041	0.06
RMSE	0.055	0.194	0.073	0.43	0.377	0.06

Experiment 4	HSAR			SAR		Multilevel
$\rho = 0.3$, $\lambda = 0.0$	ρ	λ	β_1	ρ	β_1	β_1
Bias	-0.017	-0.002	-0.024	0.286	-0.397	-0.039
SD	0.054	0.218	0.065	0.044	0.05	0.069
RMSE	0.057	0.218	0.069	0.289	0.4	0.079

Experiment 5	HSAR			SAR		Multilevel
$\rho = 0.3,\ \lambda = 0.3$	ρ	λ	β_1	ρ	β_1	β_1
Bias	-0.016	-0.075	-0.034	0.288	-0.406	-0.036
SD	0.054	0.185	0.064	0.047	0.048	0.069
RMSE	0.056	0.2	0.073	0.292	0.409	0.078

Experiment 6	HSAR			SAR		Multilevel
$\rho = 0.3,\ \lambda = 0.3$	ρ	λ	β_1	ρ	β_1	β_1
Bias	-0.017	-0.042	-0.054	0.325	-0.433	-0.041
SD	0.048	0.179	0.078	0.05	0.05	0.069
RMSE	0.051	0.184	0.095	0.329	0.436	0.08

Experiment 7	HSAR			SAR		Multilevel
$\rho = 0.5,\ \lambda = 0.0$	ρ	λ	β_1	ρ	β_1	β_1
Bias	-0.031	0.007	-0.071	0.195	-0.464	-0.141
SD	0.034	0.21	0.067	0.029	0.052	0.072
RMSE	0.046	0.21	0.097	0.197	0.467	0.158

Experiment 8	HSAR			SAR		Multilevel
$\rho = 0.5,\ \lambda = 0.3$	ρ	λ	β_1	ρ	β_1	β_1
Bias	-0.045	-0.049	-0.097	0.199	-0.476	-0.13
SD	0.04	0.18	0.068	0.03	0.051	0.066
RMSE	0.06	0.186	0.118	0.201	0.479	0.145

Experiment 9	HSAR			SAR		Multilevel
$\rho = 0.5,\ \lambda = 0.5$	ρ	λ	β_1	ρ	β_1	β_1
Bias	-0.042	-0.095	-0.118	0.219	-0.507	-0.125
SD	0.042	0.203	0.072	0.032	0.058	0.073
RMSE	0.06	0.224	0.138	0.221	0.51	0.145