The Review of Regional Studies

The Official Journal of the Southern Regional Science Association

Old and New Approaches for Spatially Varying Coefficient Models*

Dayton M. Lambert

Department of Agricultural Economics, Oklahoma State University, USA

Abstract: This note compares old and new methods for modeling spatial heterogeneity with spatially varying parameter (SVP) models. Older methods considered include spatial expansion, spatial adaptive filtering, and geographically weighted regression. Newer methods that have emerged since the beginning of the 21st include smooth transition autoregression, spatial Gaussian process, and random parameter models with autoregressive processes. A simulation is used to graphically demonstrate differences between the approaches. Regional scientists planning on using any one of these approaches should carefully consider whether the data generating process they are working with is consistent with the assumptions an SVP maintains regarding spatial heterogeneity.

Keywords: regression, parameters, heterogeneity, spatial

JEL Codes: C13, C14, R12

1. INTRODUCTION

Recent advances in computing and software have renewed interest in using local regression methods (LRM) to model spatial heterogeneity. The idea of allowing spatial units their own response to controls is conceptually appealing for regional scientists. LRM relax the assumption that all locations respond similarly to changes in local conditions, resources, or other shocks. After all, looking at a map, it is easy to conclude that migration rates into some locations is higher than some regional average, the pace of firm entry/exits exhibits localized heterogeneity, policy interventions have location-specific effects.

There is a variety of LRM available to researchers. This note uses simulation to compare graphically earlier LRM, including Casetti (1972)'s spatial expansion model, Foster and Gorr (1982)'s spatial adaptive filter (SAF), and Cleveland and Devlin (1988)'s geographically weighted regression (GWR), with more recent LRM. Recent examples of LRM include Pede et al. (2014)'s smooth transition model (STM), Finley and Banerjee (2020)'s Bayesian

* Acknowledgements. Lambert is a professor and Sparks Chair of Agricultural Sciences and Natural Resources at Oklahoma State University, Stillwater, OK 74074. E-mail: dayton.lambert@okstate.edu. This article was the Presidential Address of the 60th Southern Regional Science Association annual meeting.

Kriging with a spatial Gaussian process model (SGP), and Congdon (2020)'s spatial autoregressive random parameter model (SRP). The analysis focuses on cross-sectional data. Space-time LRM are not considered, but the results presented here can be generalized to space-time LRM. LRM are intuitively appealing and most are relatively easy to estimate, but researchers should use caution when selecting which LRM is suitable for a problem at hand. Researchers considering using LRM should consider using an exploratory approach to choose a LRM by carefully comparing the in-sample performance of each method.

2. BACKGROUND

LRM explicitly model spatial heterogeneity and not necessarily spatial autocorrelation or interaction (Cho et al. (2010)). Whittle (1954)'s introduction of spatial autoregressive processes into linear regression was an early effort to model spatial interaction between cross-sectional units. Whittle included in a linear response model the spatial lag of local outcomes as an endogenous variable. About twenty years later, Cliff and Ord (1973) reformulated Whittle's autoregressive model as one with correlated errors. Today, Whittle's, and Cliff and Ord's contributions are, what Anselin and Florax (1995) called, spatial autoregressive (SAR) or spatial error models (SEM) of the first order, respectively. Anselin (2003) summarized several variations of these models. SAR and SEM approaches model spatial interaction through auto-covariance structures, but the marginal response to local changes are global; i.e., a single estimated slope coefficient equally applies to every spatial unit. LeSage and Pace (2009) outlined procedures to decompose global effects into direct and indirect effects using the 'Leontief inverse' matrix (Anselin (2003)) implied by the SAR. Decomposed marginal responses of SAR-models provide detailed information on feedback and spillover effects, while the marginal effects of the SEM are global.

The spatial heterogeneity addressed by LRM differs qualitatively from the type of heterogeneity SAR and SEM model. For the SAR and SEM models, heterogeneity is indirectly modeled by covariance structures or, for SAR, by LeSage and Pace's decomposition for marginal effects. The heterogeneity LRM address relates to locally varying production or demand functions, which are presumably the underlying cause for spatial variation in intercept and slope coefficients (McMillen (1996)). This feature of LRM implies that slope and intercept terms for spatial units are estimable for each spatial unit.

A direct method for accommodating spatially heterogeneous slope and intercept coefficients is to assign spatial units to discrete categories that enter the response model as fixed effects (McGranahan et al. (2011)). The fixed effect approach requires external information to classify spatial units into discrete groups, for example 'north' vs. 'south', 'metropolitan' vs. 'non-metropolitan', or census divisions. An advantage of this approach is its tractability. For cross-sectional data, the fixed effect 'spatial regime' model is simple to estimate and lends well to response comparisons across regimes. A problem with this approach is that the classification scheme may not represent, or be a component of, the data generating process. This approach also maintains the assumption that spatial units clustered into mutually exclusive classifications categorical groups share the same response.

The LRM considered here are linear, cross-sectional models, all of which can be general-

ized as,

$$y_i = \sum_{k=1}^K \beta_{ik}(\mathbf{s}_i) \cdot x_{ik} + \varepsilon_i, i \in 1 \dots N, \quad \varepsilon_i \sim iid(0, \sigma^2) \quad (1)$$

with y_i an outcome observed at location i , N the number of locations in the cross-section, $\mathbf{x}_i$ a vector of controls observed at location i , $\beta_i(\mathbf{s}_i)$ an N by k vector of marginal effects, $\mathbf{s}_i$ a vector of coordinates orienting location i to its $N - 1$ neighbors, and σ^2 an error variance term. The distinguishing feature of the LRM considered here is the way in which the $\beta_i(\mathbf{s}_i)$ are parameterized.

2.1. Spatial Expansion

Casetti (1972)'s spatial expansion model specifies $\beta_i(\mathbf{s}_i)$ as a polynomial function of location coordinates $\times$ parameter interactions. For example, if a polynomial expansion of the second order is used to parameterize $\beta_i(\mathbf{s}_i)$, then,

$$\beta_{ik}(\mathbf{s}_i) = \alpha_{0k} + \alpha_{1k} \cdot xc_i + \alpha_{2k} \cdot yc_i + \alpha_{3k} \cdot xc_i^2 + \alpha_{4k} \cdot yc_i^2 + \alpha_{5k} \cdot xc_i \cdot yc_i \quad (2)$$

where the α 's are global parameters and $s_i \in (yc_i, xc_i)$ are a location's coordinates.

Higher-order polynomials are possible but at the cost of introducing collinearity. Interactions and higher order terms can be omitted to reduce problems caused by collinearity. An evident criticism of this specification is its *ad hoc* formulation pertaining to sources of heterogeneity, for which there is no apparent theoretical justification. This feature, along with collinearity introduced through the expansion terms, make inference with the spatial expansion model challenging. Anselin (2003) provides additional perspective on Casetti's spatial expansion model.

2.2. Spatial Adaptive Filter (SAF)

Foster and Gorr (1982) proposed estimating location-specific coefficients using a spatial filtering process. The method borrows from the procedures used in the time-series literature to estimate 'rolling coefficients', i.e., a moving average of coefficients that evolve over time. Spatial neighbors are analogous to the window used in time series applications that partitions a series into subsets of data.

Local coefficients are estimated as,

$$\beta_{ik}(\mathbf{s}_i) = \frac{1}{n_i} \sum_{j \in n_i} \beta_{ij}(\mathbf{s}_i) \quad (3)$$

with n_i the number of neighbors included in location i 's window. In other words, $\beta_{ik}(\mathbf{s}_i)$ is a forecast based on the weighted average of its $\beta_{ij}(\mathbf{s}_j)$ neighbors. Neighbors are defined using

continuous distance measures with a predetermined threshold or by contiguity. An iterative procedure is used to minimize the error variance of the response variable.

2.3. Geographically Weighted Regression

Cleveland and Devlin (1988) introduced what is known as geographically weighted regression (GWR). McMillen (1996) and Fotheringham et al. (2003) are generally credited for extending GWR to general linear models, including the Poisson, logit, and probit GWR estimators.

Response at location i is,

$$y_i = \sum_k \beta_{ik}(w(\mathbf{s}_i)) \cdot x_{ik} + \varepsilon_i, \quad (4)$$

where $\mathbf{s}_i$ is vector of distances between location i and its $N - 1$ neighbors, $w(\mathbf{s}_i)$ a kernel weighting function, specific to each location, and the other symbols previously defined. The kernel functions determine the scope and magnitude of location i 's neighbors' effects on its parameter estimates. Example kernels include:

$$\begin{aligned} \text{Gaussian : } w(\mathbf{s}_i) &= \phi(\rho \cdot [||\mathbf{d}_i||^2]/\hat{\sigma}_{di}) \\ \text{Exponential : } w(\mathbf{s}_i) &= \exp(-\rho \cdot ||\mathbf{d}_i||^2) \\ \text{Bi-square : } w(\mathbf{s}_i) &= (\mathbf{d}_i < d_{max}) \cdot [1 - (\mathbf{d}_i/d_{max})^2]^2 \end{aligned} \quad (5)$$

where ϕ is the standard normal probability density function, $\hat{\sigma}_{di}$ is the standard deviation of the distance vector, $\mathbf{d}_i$ a vector of distances between location i and its neighbors, and ρ a spatial parameter. The bi-square kernel is a local kernel because it works by generating subsets of observation rather than using all $N - 1$ neighbors, as the exponential and Gaussian kernels do. The bi-square parameter d_{max} determines the number of neighbors to include in the estimation of a local vector of parameters.¹ Kernel parameters are determined using cross-validation procedures (Fotheringham et al. (2003); Bivand et al. (2008)). Once the kernel parameters are recovered, GWR proceeds by estimating a vector of local parameters for each location. The matrix of GWR coefficients is,

$$\beta_i = (\mathbf{X}'\mathbf{W}(i)\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}(i)\mathbf{y} \quad (6)$$

GWR, like other LRM, is intuitively appealing and now widely available to researchers in several software environments, but it is not without problems. One issue relates to inference. As sample size increase, the number of model parameters increases at the same rate. This feature of GWR is at odds with standard asymptotic theory, which maintains that parameters converge to their true value as sample size increases. In addition, the same observations are also used repeatedly to estimate location-specific parameters. This aspect violates the standard independent and identically distributed (iid) statistical assumption regarding model

¹Note that the $(d_{ij} < d_{max})$ contrivance could also be applied to any admissible kernel function.

Figure 1: Comparison of Geographically Weighted Regression Kernels

error terms. Although it is mechanically straightforward to calculate location-specific standard errors and t-tests, it is difficult to determine what the appropriate degrees of freedom are for hypothesis testing.

Another issue with GWR relates to collinearity problems common (Farber and Páez (2007); Wheeler and Tiefelsdorf (2005)). In extreme cases, what one may interpret to be a structured geographic pattern in mapped coefficients are actually artifacts of collinearity issues arising from the GWR estimator. To address this problem, new research has extended lasso and ridge regression techniques to GWR (Wulandari et al. (2017); He et al. (2021)). These approaches are not addressed here.

A third issue pertains to the kernel choice. Unlike the practitioner's choice of a spatial weighting matrix $\mathbf{W}$ (LeSage and Pace (2009)), GWR estimates are sensitive to the kernel used to re-weight the influence of neighboring locations on location i 's estimates. Figure 1 uses simulated data to highlight the effect of kernel choice on local estimates. The model includes a constant term and a single covariate that was simulated as a standard normal random variable centered on zero with a variance of one. The error term was also a zero-

centered standard normal random variable with a variance of one, and the grid size was 40×40 , yielding 1,600 locations. The difference between the local-adaptive bi-square kernel and the Gaussian and exponential (both continuous kernels) is clear. The continuous kernels include all $N - 1$ observations as data to estimate location i 's parameters. Alternatively, fewer observations are included in the bi-square kernel estimation of local parameters. The truncated nature of the data set generated from the bi-square kernel results in a bullet pattern, whereas the patterns generated by the continuous kernel functions are smoother. These mechanical differences suggest that practitioners should carefully consider the data generating process (dgp) as it relates to local versus global random spatial fields, but also the empirical question at hand. In some cases, the dgp may be better represented as originating from localized events as opposed to a smooth, continuous, global process connecting an entire region.

For these reasons, some argue that the use of GWR, and other local regression methods, are best suited for exploratory spatial analyses (e.g., Wheeler and Calder (2007); Farber and Páez (2007)). Indeed, Roger Bivand's GWR routine for R includes this caveat (Bivand (2020)). Researchers have developed degrees-of-freedom adjustment methods for testing GWR parameter stationary and goodness-of-fit, such as Leung et al. (2000), but the approach does not appear to have been extended to location-specific hypothesis testing.

2.4. Smooth Transition Regression

Pede et al. (2014) extended regime switching, or smooth transition autoregressive models (STM) developed in the time series literature, to spatial data. The spatial STM model can accommodate SAR(1), SEM(1), and ARAR(1) spatial process models. Without loss of generality, the STM model considered here excludes spatial autoregressive processes and focuses only LRM parameter estimation.

The response equation is nonlinear in terms:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + g(\mathbf{s}; \gamma, c)\mathbf{X}\boldsymbol{\delta} + \boldsymbol{\varepsilon} \quad (7)$$

where $g(\mathbf{s}; \gamma, c)$ is a continuous cumulative density function (cdf) that sorts observations into one of two spatial regimes, $\mathbf{s}$ is a vector of distances or other location-specific data, γ is the cdf's slope, and c is a parameter identifying the geographic location of the cdf's median. The model parameters vary over locations as $\beta_{ik} = \beta_k + g(\mathbf{s}_i; \gamma, c) \cdot \delta_k$, with δ_k a slope or intercept shifter.

Pede et al. used a logistic cdf to specify the transition function:

$$g(\mathbf{s}; \gamma, c) = \frac{1}{1 + \exp(-\gamma \cdot [\mathbf{s} - c])} \quad (8)$$

Absent regimes, $\gamma = \boldsymbol{\delta} = \mathbf{0}$ and the model reduces to an aspatial linear regression. As γ increases, $g(\mathbf{s}; \gamma, c) \rightarrow 1$ and the linear model behaves as if it were a dummy variable regression, with δ_k the difference between regimes (Pede et al. (2014); Brown and Lambert (2016)). For intermediate values of γ and $\boldsymbol{\delta} \neq \mathbf{0}$, Eq. 7 is an LRM. Pede et al. restricted the slope and location parameters of the transition function to be the same across all covariates.

A straightforward extension would be to allow the slope and location parameters of $g(\cdot)$ to vary across covariates in $\mathbf{X}$ (i.e., $g(\mathbf{s}; \gamma_k, c_k)$).

2.5. Spatial Gaussian Process Models

Advances in Bayesian econometrics, and concomitantly advances in software availability and computational resources, have made it possible to model spatially varying parameters as random deviations from an overall mean response, $\bar{\beta}$, with spatially correlated deviations from the population mean. Spatially correlated deviations of a location's parameters' from the overall mean are structured as a spatial Gaussian process (SGP), usually with a semi-variogram. Both Bayesian Kriging and SAR-type SGP's can be estimated in *R* or *R-Stan* environments.

2.5.1. Bayesian Kriging and Spatial Gaussian Process

Bayesian Kriging uses resampling methods to generate posterior distributions of spatially varying estimates jointly with spatial covariance parameters. Early examples of Bayesian Kriging include Handcock and Stein (1993), Gelfand et al. (2003), and Wheeler and Calder (2007). Park et al. (2019), Finley and Banerjee (2020), and McElreath (2020) provide more recent applications. Bayesian Kriging, and Bayesian approaches more generally, are robust to degrees of freedom problems, which is a limitation of frequentist approaches that might be applied to estimate spatially varying parameter models. Nonetheless, Bayesian Kriging is computationally intensive, with the amount of time required to estimate spatially varying parameters increasingly exponentially with location units. For this reason, judiciously chosen priors aid convergence. Finley and Banerjee (2020) provide an *R* package, *spSVC*, to estimate Bayesian Kriging SGP.

The general model structure of a SGP with Bayesian Kriging is hierarchical (Finley and Banerjee (2020)):

$$\begin{aligned}
 y_i &\sim \mathcal{N}(\mu_i, \sigma) \\
 \mu_i &= \mathbf{x}_i \boldsymbol{\beta}_i \\
 p(\boldsymbol{\beta}_{ik}) &\sim \mathcal{MVN}(\bar{\boldsymbol{\beta}}_k, \mathbf{V}_k) \\
 V_{ijk} &= \begin{cases} \sigma_{\bar{\boldsymbol{\beta}}_k}^2 + \eta_{\bar{\boldsymbol{\beta}}_k}^2 \cdot g(\mathbf{d}_{ij}, \rho_k) & \text{if } i = j \\ \eta_{\bar{\boldsymbol{\beta}}_k}^2 \cdot g(\mathbf{d}_{ij}, \rho_k) & \text{if } i \neq j \end{cases} \\
 p(\bar{\boldsymbol{\beta}}) &\sim \mathcal{N}(0, 10) \\
 [p(\boldsymbol{\sigma}), p(\boldsymbol{\rho}), p(\boldsymbol{\eta})] &\sim \text{Exponential}(1)
 \end{aligned} \tag{9}$$

where y_i is a response variable that is normally distributed with a mean function of μ_i and scalar variance of σ . The mean response is linear and is a function of location-specific covariates in $\mathbf{x}_i$ with a corresponding vector of parameters, $\boldsymbol{\beta}_i$. The prior for the population mean ($\bar{\boldsymbol{\beta}}$) is normal, centered on zero with a standard deviation of 10 (a relatively diffuse

prior on variance). A reasonable prior for σ is the exponential distribution with a decay rate of one (McElreath (2020)). In other words, the average deviation of a response from the overall population mean is expected to be within ± 1 standard deviation.

The priors for the length k vector of the main effect parameters are multivariate normal with an population mean $(\bar{\beta}_k)$ with covariance structure $\mathbf{V}_k$. Spatial covariance governing the correlation of deviations between a location's effect with its $N - 1$ neighbors is modeled with a semivariogram structure, where $g(\mathbf{d}_{ij}, \rho_k)$ is exponential, Gaussian, spherical, or other admissible semivariogram functions could be used to specify $g(\cdot)$. The parameters $\sigma_{\bar{\beta}_k}^2$ and $\eta_{\bar{\beta}_k}^2$ are (respectively) nugget and sill effects of the semivariogram, and ρ_k its range (Cressie (1993)). Spatial covariance priors are also exponential with a decay rate of one.

2.5.2. SAR-type Gaussian Process Models

An interesting extension of Bayesian Kriging is Congdon (2020)'s introduction of SAR-type structures estimated as SGP. SAR-type structures of spatially varying parameters are likely more familiar to regional scientists. Congdon's approach is also a hierarchical Bayesian model. The mean response function is similar to the Bayesian Kriging SGP, in addition to the priors for the population level main effects and their standard deviations. Congdon provides *R-Stan* code to estimate a version of the SAR-type SGP.

The SAR-type SGP model is:

$$\begin{aligned}
 y_i &\sim \mathcal{N}(\mu_i, \sigma) \\
 \mu_i &= \mathbf{x}_i \boldsymbol{\beta}_i \\
 p(\beta_{ik}) &\sim \mathcal{MVN}(\beta_k, \sigma_k^2 \cdot [\mathbf{B}'_k \mathbf{B}_k]^{-1}) \\
 p(\bar{\beta}_k) &\sim \mathcal{N}(0, 10) \\
 p(\sigma), p(\sigma_k) &\sim \text{Exponential}(1) \\
 \rho_k &= 2 \cdot (z_k - 0.5) \\
 p(z_k) &\sim \text{Beta}(2, 2)
 \end{aligned} \tag{10}$$

with $\mathbf{B}_k = \mathbf{I} - \rho_k \mathbf{W}$, $\mathbf{W}$ a row-standardized spatial weights matrix, $\mathbf{I}$ is a conformable identity matrix, and $\rho_k \in (-1, 1)$ a spatial autoregressive parameter for β_{ik} . Here, the spatial autoregressive coefficient is parameterized as a function of a Beta random variable, but other priors are admissible, such as the logistic or uniform distributions.

3. SIMULATION

Knowing beforehand how these LRM estimators perform qualitatively is a first step towards selecting a suitable LRM. As one might expect, the striking differences between these LRM approaches generate very different estimates of locally varying parameters. A simulation is used to demonstrate graphically these differences. The simulation is not conducted for the GWR estimator. Unlike the other LRM examined here, there is no closed-form solution for GWR's locally varying parameters.

The simulation generates 1,600 observations uniformly spaced on a 40×40 grid. The linear model includes a constant and a single covariate with an iid error term. The dgp is:

$$\begin{aligned} y_i &= \beta_{0i} + \beta_{1i} \cdot x_i + e_i, i \in 1 \cdots 1,600 \\ \bar{\beta}_0 &= 1, \bar{\beta}_1 = 1 \\ \sigma_e^2 &= 1, \sigma_{\beta_0}^2 = 1, \sigma_{\beta_1}^2 = 1 \\ e &\sim \mathcal{N}(0, 1) \end{aligned} \quad (11)$$

The variance terms $(\sigma_{\beta_0}^2, \sigma_{\beta_1}^2)$ only pertain to the SGP and SAR-type LRM.

The locally varying intercepts and slope coefficient are calculated using the functions specific to each LRM. For example, for the SGP the β_{0i} are calculated as:

$$\beta_{0i} \sim \mathcal{MVN}(1, \mathbf{V}_0) \quad (12)$$

with

$$V_{0ij} = \begin{cases} 1 + \eta_0^2 \cdot g(\mathbf{d}_{ij}, \rho_0) & \text{if } i = j \\ \eta_0^2 \cdot g(\mathbf{d}_{ij}, \rho_0) & \text{if } i \neq j \end{cases} \quad (13)$$

The function $g(\cdot)$ is exponential. The spatial parameter ρ_0 is set to 0.01 and the sill (η_0^2) is evaluated at (1, 5). The effect of a nugget on the parameters is evaluated by removing the “1”. The location-specific slope parameters are calculated similarly.

A row-standardized Queen contiguity matrix was used for the the SAR-type LRM. Parameter fields were generated with the SAR coefficient ρ set at 0.1 (low autocorrelation) and 0.9 (high autocorrelation). Thus, for the intercept the matrix $\mathbf{V}_0$ in Eq. 12 becomes $1 \times [(\mathbf{I} - \rho_0 \mathbf{W})'(\mathbf{I} - \rho_0 \mathbf{W})]^{-1}$. The location-specific slope parameters are similarly calculated.

The spatial expansion parameters of Casetti’s model (the α ’s in Eq. 2) were set to (0, 1, 1, -0.5, -0.5, 0.5), respectively. All other model parameters were set to the values in Eq. 11. The SAF spatial weighting matrix was a first-order row standardized Queen contiguity matrix.

The STM example requires specifying a spatial variable, $\mathbf{s}_i$, for the threshold function. The simulation example uses the distance of a grid element to the center grid. The average distance of a location to the grid center is used for the location parameter c . The spatial slope and intercept shifters are $\boldsymbol{\delta} = [-3, -3]$. The shape parameter, γ , is evaluated at $\gamma \in (0, 0.25, 1, 2, 3, 5, 8, 10, 100)$, noting that at $\gamma = 0$ the STM is an aspatial linear model and at $\gamma = 100$ the STM model is a linear dummy variable regression with observations uniquely sorted into one of two groups.

4. SIMULATION RESULTS

Casetti’s spatial expansion method parameterizes local parameters as a polynomial function of coordinates. The effect of parameterizing the location-specific effects as a function of (x,

Figure 2: Simulated Field for Spatial Expansion

Note: “ y ” are predicted values.

Figure 3: Simulated Field for Spatial Adaptive Filter

Note: “ y ” are predicted values.

y)-coordinates differs for the constant and slope parameters, given the simulation parameters. A trend effect is clearly evident for the constant term since it is not dependent on location-specific data (Figure 2). The inherent randomness of the covariate dominates the observed spatial pattern, albeit with some clustering consistent with the pattern observed for the intercept term.

Compared to Casetti’s spatial expansion method, spatial clustering of intercept and slope location-specific effects are clearly evident (Figure 3). How space was defined for the SAF (neighborhood contiguity) is different than that of the spatial expansion method, which uses (x, y) -coordinates. In addition, the spatial expansion method only models surface trends, and not through Boolean contiguity with $\mathbf{W}$. An important point to note about the SAF figure is that there is no compelling reason for explaining the observed patterns, whereas with the spatial expansion method, a polynomial surface is consistent with the quadratic expansion of (x, y) -coordinate terms. For both cases of the SAF and spatial expansion, it is not immediately evident how these features constitute the dgp .

For the STM LRM, the degree of coefficient heterogeneity is governed by the transition

Figure 4: Simulated Field for STM, β_1

Notes: Spatial heterogeneity of β_1 manifests as distinct regimes.

function's slope parameter γ . The results for β_1 are examined, since similar patterns obtain for the intercept term. When $\gamma = 0$, the marginal effect of the covariate is invariant to location (Figure 4). As γ , and therefore heterogeneity, increases, the covariates's effect on the response variable becomes progressively differentiated into two distinct regimes. When $\gamma = 100$, there are effectively two discrete regimes, both with substantially different magnitudes in their effect on response. The bull's eye observed in Figure 4 is an artifact of the definition used to define spatial structure. The spatial variable $\mathbf{s}$ is a location's distance to the center of the grid. In addition, the location parameter (c) is the average over all distances, which again corresponds with the grid's center.

Figure 5: Simulated Field for Bayesian Kriging SGP Model

Note: “ y ” are predicted values.

When ρ_0 or ρ_1 is zero, then the variance term governing local parameter variation of the SGP is compound with $\text{var}(\beta_{ik}) = \eta_{\beta_k}^2 + \sigma_{\beta_k}^2$. In other words, the distribution of location effects is random and unstructured, as evidenced by Figure 5 (first row). The effect of including a nugget is also apparent (second and third rows, Figure 5). Inclusion of a nugget effect adds a random component through σ . Omitting a nugget in the SGP specification tends to smooth, and potentially oversmooth, the parameter’s surface.

The SAR-type LRM parameterization of location-specific effects is similar to that of the Bayesian Kriging SGP considered here, but for the structure of the covariance process. At low levels of spatial autocorrelation, the distribution of location-specific parameters appears to be random (Figure 6). As correlation between locations increases, so too does the spatial structure exhibited by the parameters.

Figure 6: Simulated Field for Spatial Random Parameter Regression

Note: “ y ” are predicted values.

5. CONCLUSIONS

Regional scientists are in a unique position to contribute to the advancement of spatially varying parameter models, given the spatially inherent nature of the data we use. The objective of this demonstration was to provide a qualitative comparison of different LRM procedures, while attempting to minimize differences attributable to scale, complexity, and other model artifacts. Each LRM has its own strengths and weaknesses. The differences in the LRM considered here are largely due to the way in which spatial process is defined and structured.

From the standpoint of computational efficiency, Casetti’s spatial expansion and GWR are superior compared to the other LRM estimators considered here. The *ad hoc* specifications of the spatial expansion and SAF estimators makes it difficult to interpret patterns of parameter heterogeneity revealed by these approaches. The spatial expansion procedure models parameter variation as a polynomial trend. If the actual dgp is unrelated to the expansion function, then the empirical model may be grossly misspecified. The same caveat applies to the SAF LRM.

If theory suggests that certain control variables cause structural breaks across space, then the STM model is consistent with this notion. In the contrived example provided here, the spatial variable included in the regime function was designed to emphasize regime switching potential. In practice, one would include the spatial lag of a covariate in the regime splitting function (Pede et al. (2014)), or the distance of a location to important features, such as

metropolitan counties. Alternatively, levels of covariates can be included in the function $g(\cdot)$, such as population density.

The most computationally complicated LRM are the SGP and SAR-type estimators. Like GWR, these two estimators require specification of a covariance process. The decision of whether one should include a nugget effect in the SGP's covariance term is nontrivial, as evidenced by the intensity of its smoothing effects on parameter heterogeneity. Omission of a nugget effect may lead to parameter over-smoothing. Conversely, the SAR-type LRM indirectly includes nugget effects through the parameter's covariance terms. Casetti's spatial expansion model and Gorr and Foster's SAF were first to suggest an approach to model heterogeneity with spatially varying parameters. Spatial expansion models can be estimated using ordinary least squares. However, introduction of polynomial expansion terms introduces collinearity. In addition, the expansion procedure may, or may not, be related to the *dgp* faced by the researcher. The SAF is more difficult to estimate and requires an optimization procedure to tune separately each location's parameter vector. Like the spatial expansion method, the SAF may be disconnected with the true data generating process.

GWR is a popular alternative for modeling heterogeneity with spatially varying parameters. Unlike the spatial expansion and SAF approaches, GWR does not rely on *ad hoc* specifications to estimate spatially varying parameters. Current algorithms are efficient in terms of estimating GWR models using large spatial data sets. However, rather than using default options in public domain software, researchers should consider carefully the type of kernel used to model heterogeneity with GWR, and how kernel choice informs the presumed *dgp*.

The STM procedure is another alternative for modeling heterogeneity as a spatially varying parameter issue. An advantage of the STM is that it provides a method to test for structural breaks in parameter variation across space. This feature may be of interest to researchers studying the role of agglomeration economies in regional growth dynamics. Future research could consider revising the regime function to one that is associated with each covariate included in the regression model.

Advances in software availability, increased access to open source repositories, and computing power have popularized hierarchical Bayesian modeling approaches for addressing complicated covariance structures and model design. These advances have opened the door to modeling complicated spatial covariance structures in addition to spatially varying parameters. However, like GWR, care must be given to kernel selection, and, in the case of Bayesian Kriging models, the components of the semivariograms used to model parameter covariance. In addition, although computational advances make estimation of spatial Bayesian hierarchical models tractable, larger spatial data sets continue to pose computational challenges in terms of convergence time and parameter stability.

REFERENCES

- Anselin, Luc. (2003) *Spatial Econometrics*. Kluwer Academic Publishers.
- Anselin, Luc and Raymond JGM Florax. (1995) "Small sample properties of tests for spatial dependence in regression models: Some further results," In *New Directions in Spatial Econometrics*. Springer: pp. 21–74.
- Bivand, Roger. (2020) "Geographically Weighted Regression," <https://cran.r-project.org/web/packages/spgwr/vignettes/GWR.pdf>.
- Bivand, Roger S, Edzer J Pebesma, Virgilio Gómez-Rubio, and Edzer Jan Pebesma. (2008) *Applied Spatial Data Analysis with R*, volume 747248717. Springer.
- Brown, Jason P and Dayton M Lambert. (2016) "Extending a Smooth Parameter Model to Firm Location Analyses: The Case of Natural Gas Establishments in the United States," *Journal of Regional Science*, 56(5), 848–867. <http://doi.org/10.1111/jors.12280>.
- Casetti, Emilio. (1972) "Generating Models by the Expansion Method: Applications to Geographical Research," *Geographical analysis*, 4(1), 81–91. <http://doi.org/10.1111/j.1538-4632.1972.tb00458.x>.
- Cho, Seong-Hoon, Dayton M Lambert, and Zhuo Chen. (2010) "Geographically Weighted Regression Bandwidth Selection and Spatial Autocorrelation: An Empirical Example Using Chinese Agriculture Data," *Applied Economics Letters*, 17(8), 767–772. <http://doi.org/10.1080/13504850802314452>.
- Cleveland, William S and Susan J Devlin. (1988) "Locally Weighted Regression: An Approach to Regression Analysis by Local Fitting," *Journal of the American Statistical Association*, 83(403), 596–610. <http://doi.org/10.1080/01621459.1988.10478639>.
- Cliff, Andrew D. and J.K. Ord. (1973) "Spatial Autocorrelation," *Progress in Human Geography*, 19(2), 245–249. <http://doi.org/10.1177/030913259501900205>.
- Congdon, Peter D.. (2020) *Bayesian Hierarchical Models*. Chapman and Hall/CRC.
- Cressie, Noel A. C. (1993). *Statistics for Spatial Data*, pp. 1–26. John Wiley & Sons, Ltd.
- Farber, Steven and Antonio Páez. (2007) "A Systematic Investigation of Cross-Validation in GWR Model Estimation: Empirical Analysis and Monte Carlo Simulations," *Journal of Geographical Systems*, 9(4), 371–396. <http://doi.org/10.1007/s10109-007-0051-3>.
- Finley, Andrew O and Sudipto Banerjee. (2020) "Bayesian Spatially Varying Coefficient Models in the SpBayes R Package," *Environmental Modelling & Software*, 125, 104608. <http://doi.org/10.1016/j.envsoft.2019.104608>.
- Foster, Stuart A and Wilpen L Gorr. (1982) "Federal Health Care Policy and the Geographic Diffusion of Physicians: A Macro-Scale Analysis," *Policy Sciences*, 25(2), 117–134. <http://doi.org/10.1007/BF00233744>.
- Fotheringham, A Stewart, Chris Brunsdon, and Martin Charlton. (2003) *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships*. John Wiley & Sons.
- Gelfand, Alan E, Hyon-Jung Kim, CF Sirmans, and Sudipto Banerjee. (2003) "Spatial Modeling with Spatially Varying Coefficient Processes," *Journal of the American Statistical Association*, 98(462), 387–396. <http://doi.org/10.1198/016214503000170>.
- Handcock, Mark S and Michael L Stein. (1993) "A Bayesian Analysis of Kriging," *Technometrics*, 35(4), 403–410. <http://doi.org/10.1080/00401706.1993.10485354>.
- He, Yuxin, Yang Zhao, and Kwok Leung Tsui. (2021) "An Adapted Geographically Weighted

- LASSO (Ada-GWL) Model for Predicting Subway Ridership,” *Transportation*, 48(3). <http://doi.org/10.1007/s11116-020-10091-2>.
- LeSage, James and Robert Kelley Pace. (2009) *Introduction to Spatial Econometrics*. Chapman and Hall, CRC Press.
- Leung, Yee, Chang-Lin Mei, and Wen-Xiu Zhang. (2000) “Statistical Tests for Spatial Non-stationarity Based on the Geographically Weighted Regression Model,” *Environment and Planning A*, 32(1), 9–32. <http://doi.org/10.1068/a3162>.
- McElreath, R.. (2020) *Statistical Rethinking*. CRC Press.
- McGranahan, David A, Timothy R Wojan, and Dayton M Lambert. (2011) “The Rural Growth Trifecta: Outdoor Amenities, Creative Class and Entrepreneurial Context,” *Journal of Economic Geography*, 11(3), 529–557. <http://doi.org/10.1093/jeg/lbq007>.
- McMillen, Daniel P. (1996) “One Hundred Fifty Years of Land Values in Chicago: A Non-parametric Approach,” *Journal of Urban Economics*, 40(1), 100–124. <http://doi.org/10.1006/juec.1996.0025>.
- Park, Eun Chun, B Wade Brorsen, and Ardian Harri. (2019) “Using Bayesian kriging for Spatial Smoothing in Crop Insurance Rating,” *American Journal of Agricultural Economics*, 101(1), 330–351. <http://doi.org/10.1093/ajae/aay045>.
- Pede, Valerien O, Raymond JGM Florax, and Dayton M Lambert. (2014) “Spatial Econometric STAR models: Lagrange Multiplier Tests, Monte Carlo simulations and an Empirical Application,” *Regional Science and Urban Economics*, 49, 118–128. <http://doi.org/10.1016/j.regsciurbeco.2014.07.001>.
- Wheeler, David and Michael Tiefelsdorf. (2005) “Multicollinearity and Correlation Among Local Regression Coefficients in Geographically Weighted Regression,” *Journal of Geographical Systems*, 7(2), 161–187. <http://doi.org/10.1007/s10109-005-0155-6>.
- Wheeler, David C and Catherine A Calder. (2007) “An Assessment of Coefficient Accuracy in Linear Regression Models with Spatially Varying Coefficients,” *Journal of Geographical Systems*, 9(2), 145–166. <http://doi.org/10.1007/s10109-006-0040-y>.
- Whittle, P.. (1954) “On Stationary Processes in the Plane,” *Biometrika*, 41(3-4), 434–449. <http://doi.org/10.1093/biomet/41.3-4.434>.
- Wulandari, Reny, R. A. Saefuddin, and F. M. Afendi. (2017) “Application Geographically Weighted Ridge Regression and Geographically Weighted Lasso in Case Correlation of Covariate,” *International Journal of Scientific & Engineering Research*, 8, 445–448.