This is the pre-peer-reviewed version of "Estimation and Prediction in the Presence of Spatial Confounding For Spatial Linear Models" Scandinavian Journal of Statistics 44(3) pages 780-797 which has been published in final form at onlinelibrary.wiley.com/doi/10.1111/sjos.12275/full

Estimation and Prediction in the Presence of Spatial Confounding for Spatial Linear Models

Garritt L. Page
Department of Statistics
Brigham Young University
page@stat.byu.edu

Yajun Liu
Wells Fargo Bank
Chapel Hill, North Carolina
junuylia@hotmail.com

Zhuoqiong He
Department of Statistics
University of Missouri
hezh@missouri.edu

Dongchu Sun
Department of Statistics
University of Missouri
East Chinese Normal University
sund@missouri.edu

February 17, 2017

Abstract

In studies that produce data with spatial structure it is common that covariates of interest vary spatially in addition to the error. Because of this, the error and covariate are often correlated. When this occurs it is difficult to distinguish the covariate effect from residual spatial variation. In an *iid* normal error setting, it is well known that this type of correlation produces biased coefficient estimates but predictions remain unbiased. In a spatial setting recent studies have shown that coefficient estimates remain biased, but spatial prediction has not been addressed. The purpose of this paper is to provide a more detailed study of coefficient estimation from spatial models when covariate and error are correlated and then begin a formal study regarding spatial prediction. This is carried out by investigating properties of the generalized least squares estimator and the best linear unbiased predictor when a spatial random effect and a covariate are jointly modeled. Under this setup we demonstrate that the mean squared prediction error is possibly reduced when covariate and error are correlated.

Keywords: confounding bias; generalized least squares estimator; spatial prediction; spatial correlation.

1 Introduction

Many epidemiological, ecological, and geological studies (among others) are spatial in nature in that observations taken from experimental units or subjects display some type of spatial correlation. It is common for these types of studies to be conducted with the purpose of determining relationships between covariates and response (e.g., the proximity to a water source and cancer rates) and/or make predictions. Spatial regression models are typically employed to model data that have been collected for these types of study objectives and spatial structure is commonly incorporated by introducing latent experimental unit (or subject) random effects. Because covariates of interest are measured on each experimental unit, they also may exhibit spatial structure and as a result are often correlated with the random effect. When this occurs it is not easy to separate the regression effect from that of spatially varying error since they are no longer orthogonal. This inability to separate covariate effects from spatial error is referred to as spatial confounding.

Spatial confounding has been recognized for some time. Clayton et al. (1993) is probably the first article to highlight spatial confounding, but since then Woodard et al. (1999), Reich et al. (2006), and Wakefield (2007) (among others) have in one way or another noted that incorporating spatial residual structure in regression models can produce puzzling results. Formal studies of spatial confounding and its influence on inference have appeared more recently. Paciorek (2010) considered spatial scale's influence on biases induced by spatial confounding. He along with Hodges and Reich (2010) showed that the practice of including a spatially varying random effect does not necessarily reduce omitted variable bias. Further, Hodges and Reich (2010) argued that spatial confounding is ubiquitous in spatial regression and is not restricted to a particular type of model, but possibly may be present in any regression model that incorporates the notion that units near each other produce measurements more similar than units further apart.

A few solutions to avoiding spatial confounding's effect on covariate inference have also recently appeared in the literature. See for example Hodges and Reich (2010), Hughes and Haran (2013), Caragea and Kaiser (2009), and Lee et al. (2014). These solutions address dependence between covariate and spatial random effect by orthogonalizing in one way or another the column spaces associated with each one resulting in the so called restricted spatial regression (RSR) model. Recently, Hanks et al. (2015) considered the utility of the RSR type models through an extensive simulation study. They showed that an increased effective range resulted in an increased Type-S error for coefficient estimates produced by the RSR model. (For Hanks et al. (2015) a Type-S error occurred if an interval associated with a parameter that is zero does not contain zero. See Gelman and Tuerlinckx 2000 for more details.) Paciorek (2010) also showed that effective range is an important factor in spatial confounding bias. We briefly note that Paciorek (2010) used the term "spatial scale" in place of "spatial range". This terminology may be a bit confusing ("spatial scale" can also refer to the fineness or proximity of spatial units), but in what follows we still use them interchangeably and clarify when necessary.

Even though work dedicated to studying spatial confounding bias is present in the literature, to our knowledge, the focus has been on coefficient estimation and no formal study regarding spatial confounding's influence on spatial prediction exists (Hanks et al. (2015) mention prediction in the context of partially observed fields when studying RSR, but do not formally consider it). A possible reason for this is that under an *iid* regression, predictions that don't require extrapolation remain unbiased even in the presence of collinearity and/or dependence between error and covariate. However, it is not obvious that this is the case in a spatial setting and the principal purpose of this paper is to begin a formal study of spatial confounding's impact on prediction. To this end we first consider covariate coefficient estimation under the same basic set up found in Paciorek (2010) and further knowledge regarding how spatial confounding affects the generalized least squared estimator (GLS) when correlation between a random effect and covariate is ignored. Once a general framework has been established, we study properties of the the kriging predictor in the presence of spatial confounding. It turns out that if correlation between covariate and error is ignored when in fact they are jointly normal, then the mean squared prediction error (MSPE) of the kriging predictor can, in some instances, be reduced.

We briefly mention that what we call spatial confounding is very similar to what econometricians and social scientists call endogeneity in a non-spatial setting. Because the presence of endogenous variables is common in those fields much work has been dedicated to develop methods that handle them. However, just as in the statistical literature, to our knowledge no formal study regarding prediction and endogeneity has been carried out. Also, there is a literature that considers the bias of regression coefficients in a spatial setting when random effects are considered as formal devices to implement some type of smoothing (Green 1985, Speckman 1988, and Besag and Higdon 1999). Hodges and Reich (2010) and Paciorek (2010) also address this perspective showing that spatial confounding persists.

The remainder of the article is organized as follows. In Section 2 we focus on analytical results for covariate estimation under a joint normal model for random effects and covariate. In Section 3, we turn our attention to prediction and under the same joint model consider bias and MSPE of the kriging predictor. Some concluding remarks are provided in Section 4. Note that an online supplementary file that contains additional numerical results accompanies this article.

2 Estimation of a Single Covariate

We focus on analytical results that are tractable when the error structure is completely known. Although this setting is unrealistic in most applied contexts, it does provide a foundation on which investigating characteristics of the GLS estimator under the proposed model can begin. We follow the approach of Paciorek (2010) and consider covariates as being random (which is reasonable in most regression settings and defensible in some experimental design settings). We derive the expected value, bias, and MSE of the GLS estimator from a spatial regression setting when a measured covariate varies spatially and is correlated with a spatially varying error.

2.1 Setup of Theoretical Framework

Consider a simpler linear mixed regression model with spatial structure,

$$\boldsymbol{y} = \beta_0 \mathbf{1}_n + \beta_1 \boldsymbol{x} + \boldsymbol{z} + \boldsymbol{e}, \quad \boldsymbol{e} \sim N_n(\mathbf{0}, \sigma^2 \boldsymbol{I}_n), \tag{1}$$

where $\boldsymbol{x}$ is a covariate and $\boldsymbol{z}$ is a spatially varying random effect. Here, $\boldsymbol{y}, \boldsymbol{x}, \boldsymbol{z}$, and $\boldsymbol{e}$ are all location dependent n-dimensional vectors, and $\mathbf{1}_n$ is a n-dimensional vector of 1's. Spatial structure is often induced in the model by assuming $\boldsymbol{z} \sim N(\mathbf{0}, \sigma_z^2 \boldsymbol{R}_z(\boldsymbol{\theta}_z))$, where $\boldsymbol{R}_z(\boldsymbol{\theta}_z)$ is a correlation matrix whose spatial structure is parametrized by $\boldsymbol{\theta}_z$. In model (1), $\boldsymbol{x}$ and $\boldsymbol{z}$ are assumed to be independent (at least implicitly). Under this assumption it is possible to marginalize $\boldsymbol{y}$ over $\boldsymbol{z}$ to get

$$\boldsymbol{y} \mid \boldsymbol{x} \sim N(\beta_0 \mathbf{1}_n + \beta_1 \boldsymbol{x}, \sigma^2 \boldsymbol{\Sigma}), \; \boldsymbol{\Sigma} = \boldsymbol{I}_n + \eta \boldsymbol{R}_z(\boldsymbol{\theta}_z), \tag{2}$$

where $\eta = \sigma_z^2/\sigma^2$. More succinctly, if we define $\boldsymbol{X} = (\mathbf{1}_n, \boldsymbol{x})$ and $\boldsymbol{\beta} = (\beta_0, \beta_1)'$, (2) can be written as $\boldsymbol{y} \mid \boldsymbol{x} \sim N(\boldsymbol{X}\boldsymbol{\beta}, \sigma^2 \boldsymbol{\Sigma})$.

$\boldsymbol{R}_z(\boldsymbol{\theta}_z)$ can take on a variety of forms, the selection of which depends on the data types and modeling goals. For example, if point-referenced or geostatistical data are available, a common structure is $\boldsymbol{R}_z(\boldsymbol{\theta}_z) = (\text{Corr}(\boldsymbol{s}_i, \boldsymbol{s}_j; \boldsymbol{\theta}_z))$, where $\text{Corr}(\boldsymbol{s}_i, \boldsymbol{s}_j; \boldsymbol{\theta}_z)$ is a valid correlation function with $(\boldsymbol{s}_1, \ldots, \boldsymbol{s}_n) \in \mathcal{D} = \mathbb{R}^l$ and spatial structure parameterized by $\boldsymbol{\theta}_z$. If areal data are available, then a conditional autoregressive (CAR) model can be used to model $\boldsymbol{z}$. One specification of a CAR model is for $\boldsymbol{\theta}_z = \kappa_z$ and $\boldsymbol{R}_z(\kappa_z) = (\boldsymbol{I}_n - \kappa_z \boldsymbol{C})^{-1}$ for $\kappa_z \in (\lambda_{min}^{-1}, \lambda_{max}^{-1})$. $\boldsymbol{C}$ is an adjacency matrix, λ_{min}, λ_{max} are the smallest and largest eigenvalues of $\boldsymbol{C}$ and κ_z is parameter associated with spatial dependence. The restriction on κ_z ensures that $(\boldsymbol{I}_n - \kappa_z \boldsymbol{C})^{-1}$ is positive definite and nonsingular. We list just two possible forms for $\boldsymbol{R}_z(\boldsymbol{\theta}_z)$ many others are available (see Cressie 1993 and Banerjee et al. 2014).

If $\boldsymbol{\Sigma}$ is known, a consistent and efficient estimator for $\boldsymbol{\beta}$ is the GLS estimator given by:

$$\hat{\boldsymbol{\beta}}^G = \boldsymbol{A}\boldsymbol{y}, \text{ where } \boldsymbol{A} = (\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}. \tag{3}$$

(Note that $\boldsymbol{A}$ is a function of η.) The GLS estimator $\hat{\boldsymbol{\beta}}^G$ is also the maximum likelihood estimator under (2). From a Bayesian perspective, with likelihood found in (2) and $\pi(\boldsymbol{\beta}) \propto 1$ as a prior distribution, the posterior distribution of $\boldsymbol{\beta}$ is $\boldsymbol{\beta} \mid \boldsymbol{y}, \boldsymbol{x}, \sigma^2 \sim N_2(\boldsymbol{A}\boldsymbol{y}, \sigma^2(\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X})^{-1})$ and so $\hat{\boldsymbol{\beta}}^G = E(\boldsymbol{\beta} \mid \boldsymbol{y}, \boldsymbol{x}, \sigma^2)$. Thus, $\hat{\boldsymbol{\beta}}^G$ is a fairly universal estimator for $\boldsymbol{\beta}$. Now through well known multivariate normal properties, the sampling distribution of $\hat{\boldsymbol{\beta}}^G$ given $\boldsymbol{x}$ is

$$(\hat{\boldsymbol{\beta}}^G \mid \boldsymbol{\beta}, \sigma_z^2, \boldsymbol{\theta}_z, \sigma^2, \boldsymbol{x}) \sim N_2(\boldsymbol{\beta},\ \sigma^2(\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X})^{-1}). \tag{4}$$

In addition to the distributional result, an immediate consequence of (4) is that $\hat{\boldsymbol{\beta}}^G$ is an unbiased estimator of $\boldsymbol{\beta}$ with variance $\sigma^2(\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X})^{-1}$.

As mentioned previously, (4) is available when $\boldsymbol{x}$ and $\boldsymbol{z}$ are independent. However, if $\boldsymbol{x}$ is correlated with $\boldsymbol{z}$ (e.g., $\boldsymbol{x} \sim N_n(\boldsymbol{0}, \sigma_x^2 \boldsymbol{R}_x(\boldsymbol{\theta}_x)$ and $\text{Cov}(\boldsymbol{x}, \boldsymbol{z}) = \rho\sigma_x\sigma_z \boldsymbol{R}_x^{1/2}(\boldsymbol{\theta}_x)\boldsymbol{R}_z^{1/2'}(\boldsymbol{\theta}_z))$, then $\hat{\boldsymbol{\beta}}^G$ is no longer an unbiased estimator of $\boldsymbol{\beta}$ and the covariance matrix of $\hat{\boldsymbol{\beta}}^G$ is changed. This is due to the fact that the GLS estimator ignores any correlation between the residual and the covariate. In fact, as noted in Paciorek (2010), if $\boldsymbol{x}$ is correlated with $\boldsymbol{z}$ then it would not be appropriate to marginalize $\boldsymbol{y}$ over $\boldsymbol{z}$ to obtain (2). In a regression setting where $(\boldsymbol{y} \mid \boldsymbol{x})$ is of principal interest it would be more appropriate to marginalize $\boldsymbol{y}$ over $(\boldsymbol{z} \mid \boldsymbol{x})$. We develop this concept further for the case when $\boldsymbol{x}$ and $\boldsymbol{z}$ are jointly normal in the next section.

2.1.1 Analytical Results of the Generalized Least Squares Estimator for $\boldsymbol{\beta}$

Consider the case where $\boldsymbol{x}$ and $\boldsymbol{z}$ are jointly normal with the following structure

$$\begin{pmatrix} \boldsymbol{x} \\ \boldsymbol{z} \end{pmatrix} \sim N_{2n}\left(\begin{bmatrix} \boldsymbol{\mu}_x \\ \boldsymbol{\mu}_z \end{bmatrix}, \begin{bmatrix} \sigma_x^2 \boldsymbol{R}_x & \rho\sigma_x\sigma_z \boldsymbol{R}_x^{1/2}\boldsymbol{R}_z^{1/2'} \\ \rho\sigma_x\sigma_z \boldsymbol{R}_z^{1/2}\boldsymbol{R}_x^{1/2'} & \sigma_z^2 \boldsymbol{R}_z \end{bmatrix} \right), \tag{5}$$

where $\boldsymbol{R}_z^{1/2}$ is a $n \times n$ matrix satisfying $\boldsymbol{R}_z = \boldsymbol{R}_z^{1/2}\boldsymbol{R}_z^{1/2'}$ and may be a Cholesky decomposition or an eigenvalue decomposition. Here and in what follows for notational convenience we suppress denoting $\boldsymbol{R}_x$ and $\boldsymbol{R}_z$ explicitly as a function of $\boldsymbol{\theta}_x$ and $\boldsymbol{\theta}_z$ respectively. Using well known properties of the normal distribution, under (5) the distribution of $(\boldsymbol{z} \mid \boldsymbol{x})$ is

$$(\boldsymbol{z} \mid \boldsymbol{x}) \sim N_n\big(\boldsymbol{\mu}_z + \rho\frac{\sigma_z}{\sigma_x}\boldsymbol{R}_z^{1/2}\boldsymbol{R}_x^{-1/2}(\boldsymbol{x} - \boldsymbol{\mu}_x),\ \sigma_z^2(1-\rho^2)\boldsymbol{R}_z\big). \tag{6}$$

Paciorek (2010) begins his study with the same cross-covariance structure as in (5) but with the added condition that $\boldsymbol{R}_x = \boldsymbol{R}_z$ resulting in a simple separable cross-covariance model (see Section 9.3 of Banerjee et al. 2014). What we propose in (5) is a type of coregionalization model (see Section 9.5 of Banerjee et al. 2014) that allows $\boldsymbol{x}$ and $\boldsymbol{z}$ to be completely different stochastic processes. Additionally, Paciorek (2010) assumes $\boldsymbol{\mu}_z = \boldsymbol{\mu}_x = \boldsymbol{0}$ an assumption that we adopt as $\boldsymbol{\mu}_z = \boldsymbol{0}$ is a common assumption in random effects models and $\boldsymbol{\mu}_x = \boldsymbol{0}$ is analogous to centering $\boldsymbol{x}$ (with out loss of generality). Now, under (5) and $\boldsymbol{\mu}_x = \boldsymbol{\mu}_z = 0$, the conditional distribution of the response given the covariate $(\boldsymbol{y} \mid \boldsymbol{x})$ is

$$(\boldsymbol{y} \mid \boldsymbol{x}) \sim N_n\big(\boldsymbol{X\beta} + \rho\frac{\sigma_z}{\sigma_x}\boldsymbol{R}_z^{1/2}\boldsymbol{R}_x^{-1/2}\boldsymbol{x},\ \sigma^2\boldsymbol{\Sigma}_\rho\big), \tag{7}$$

where $\boldsymbol{\Sigma}_\rho = \boldsymbol{I}_n + \eta(1-\rho^2)\boldsymbol{R}_z$. This result will be used in Section 3. We briefly note that the same joint modeling is employed in the simulation studies found in Hanks et al. (2015), except they assume that $\rho = 0$. To investigate changes to the sampling distribution of $(\hat{\boldsymbol{\beta}}^G \mid \boldsymbol{x})$ if correlation between $\boldsymbol{x}$ and $\boldsymbol{z}$ is ignored, we derive a few statistical properties of $\hat{\boldsymbol{\beta}}^G$ under (5).

Proposition 1. *If $\boldsymbol{x}$ and $\boldsymbol{z}$ are jointly distributed as in* (5) *with all variance components known, then the sampling distribution of $\hat{\boldsymbol{\beta}}^G$ defined in* (3) *is*

$$(\hat{\boldsymbol{\beta}}^G \mid \boldsymbol{\beta}, \sigma_z^2, \sigma_x^2, \sigma^2, \boldsymbol{\theta}_x, \boldsymbol{\theta}_z, \rho, \boldsymbol{x}) \sim N_2\big(\boldsymbol{\beta} + \rho\frac{\sigma_z}{\sigma_x}\boldsymbol{A}\boldsymbol{R}_z^{1/2}\boldsymbol{R}_x^{-1/2}\boldsymbol{x},\ \sigma^2\boldsymbol{\Sigma}^G\big), \tag{8}$$

where $\boldsymbol{\Sigma}^G = (\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X})^{-1} - \rho^2\eta\boldsymbol{A}\boldsymbol{R}_z\boldsymbol{A}'$ *and* $\eta = \sigma_z^2/\sigma^2$.

Proof. See the Appendix □

Remark 1. Under (5), $\hat{\boldsymbol{\beta}}^G$ is now a biased estimator of $\boldsymbol{\beta}$ with

$$\text{bias}(\hat{\boldsymbol{\beta}}^G \mid \boldsymbol{x}, \rho) = \rho\frac{\sigma_z}{\sigma_x}\boldsymbol{A}\boldsymbol{R}_z^{1/2}\boldsymbol{R}_x^{-1/2}\boldsymbol{x}. \tag{9}$$

As expected, $\text{bias}(\hat{\boldsymbol{\beta}}^G \mid \boldsymbol{x}, \rho) \to \mathbf{0}$ as $\rho \to 0$. Also, because $\boldsymbol{A}\boldsymbol{x} = (0, 1)'$, under $\boldsymbol{R}_z = \boldsymbol{R}_x$

$$\text{bias}(\hat{\boldsymbol{\beta}}^G \mid \boldsymbol{x}, \rho) = \rho\frac{\sigma_z}{\sigma_x}\boldsymbol{A}\boldsymbol{x} = \left(0, \rho\frac{\sigma_z}{\sigma_x}\right)'. \tag{10}$$

Thus, when spatial structure of $\boldsymbol{z}$ is equal to that of $\boldsymbol{x}$, the bias incurred in the GLS estimator of β_1 is simply the regression coefficient that results when $\boldsymbol{x}$ is regressed on $\boldsymbol{z}$ which is a result found in Paciorek (2010).

Remark 2. Regarding the variance we highlight two results. First, under (5) the variance of $\hat{\boldsymbol{\beta}}^G$ is always *smaller* compared to when $\boldsymbol{x}$ and $\boldsymbol{z}$ are independent. Secondly, $\text{Var}(\hat{\boldsymbol{\beta}}^G \mid \boldsymbol{x}, \rho)$ does not depend explicitly on the spatial structure or the variability of $\boldsymbol{x}$ $(\sigma_x^2\boldsymbol{R}_x)$.

Since the MSE incorporates both bias and variance, we turn our attention to its calculation in the next proposition.

Proposition 2. *Under the assumptions of proposition 1 the* MSE *of* $\hat{\boldsymbol{\beta}}^G$ *is*

$$\text{MSE}(\hat{\boldsymbol{\beta}}^G \mid \boldsymbol{x}, \rho) = \rho^2 \frac{\sigma_z^2}{\sigma_x^2} \operatorname{tr}\{\boldsymbol{A}\boldsymbol{K}\boldsymbol{A}'\} + \sigma^2 \operatorname{tr}\{(\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X})^{-1}\} - \rho^2\sigma_z^2 \operatorname{tr}\{\boldsymbol{A}\boldsymbol{R}_z\boldsymbol{A}'\}, \tag{11}$$

where $\boldsymbol{K} = \boldsymbol{R}_z^{1/2}\boldsymbol{R}_x^{-1/2}\boldsymbol{x}\boldsymbol{x}'\boldsymbol{R}_x^{-1/2'}\boldsymbol{R}_z^{1/2'}$. *Note that the first term on the right hand side of (11) is* $[\text{bias}(\hat{\boldsymbol{\beta}}^G \mid \boldsymbol{x}, \rho)]'[\text{bias}(\hat{\boldsymbol{\beta}}^G \mid \boldsymbol{x}, \rho)]$ *and that the second term is* $\text{MSE}(\hat{\boldsymbol{\beta}}^G \mid \boldsymbol{x}, \rho = 0)$. *So,*

$$\text{MSE}(\hat{\boldsymbol{\beta}}^G \mid \boldsymbol{x}, \rho) - \text{MSE}(\hat{\boldsymbol{\beta}}^G \mid \boldsymbol{x}, \rho = 0) = \rho^2 \frac{\sigma_z^2}{\sigma_x^2} \operatorname{tr}\{\boldsymbol{A}(\boldsymbol{K} - \boldsymbol{R}_z)\boldsymbol{A}'\}. \tag{12}$$

The form of (12) indicates that the slope of the regression of x onto z influences the MSE. That said, the form of (12) makes it difficult to develop intuition regarding how σ_z^2, σ_x^2 and σ^2 interact to influence the MSE as $\boldsymbol{A}$ is a complex function of $\eta = \sigma_z^2/\sigma^2$. In fact, it is not even obvious that $\boldsymbol{A}(\boldsymbol{K} - \boldsymbol{R}_z)\boldsymbol{A}'$ is positive definite. Knowing the positive definiteness $\boldsymbol{A}(\boldsymbol{K} - \boldsymbol{R}_z)\boldsymbol{A}'$ would aid in determining if (12) is positive or negative since the trace of a matrix is the sum of its eigenvalues and the eigenvalues of a positive definite matrix are all positive. Even if $\boldsymbol{R}_x = \boldsymbol{R}_z = \boldsymbol{R}$ is assumed and

$$\begin{aligned} \text{MSE}(\hat{\boldsymbol{\beta}}^G \mid \boldsymbol{x}, \rho) - \text{MSE}(\hat{\boldsymbol{\beta}}^G \mid \boldsymbol{x}, \rho = 0) &= \rho^2\sigma_z^2 \operatorname{tr}\{\boldsymbol{A}\Big(\frac{1}{\sigma_x^2}\boldsymbol{x}\boldsymbol{x}' - \boldsymbol{R}\Big)\boldsymbol{A}'\} \\ &= \rho^2 \frac{\sigma_z^2}{\sigma_x^2} - \rho^2\sigma_z^2 \operatorname{tr}\{\boldsymbol{A}\boldsymbol{R}\boldsymbol{A}'\} \end{aligned} \tag{13}$$

it is still not clear how the MSE of $\hat{\boldsymbol{\beta}}^G$ is influenced by the variance components of (1) and (5). In light of this, we resort to numerically exploring the MSE and bias of $\hat{\boldsymbol{\beta}}^G$.

2.2 Numerical Exploration

To numerically explore how the covariance parameters in (1) and (5) influence the MSE and bias of $\hat{\beta}_1^G$ when correlation between $\boldsymbol{x}$ and $\boldsymbol{z}$ is ignored, we take on the areal data modeling scenario of He and Sun (2000) who estimated county level hunting success rates using postseason turkey harvest surveys for the state of Missouri. A simultaneous CAR model structure (Clayton and Kaldor 1987) was used and thus $\boldsymbol{R}_z = (\boldsymbol{I} - \kappa_z \boldsymbol{C})^{-1}$, where $\boldsymbol{C} = (C_{kl})$ is a 114×114 symmetric adjacency matrix such that $C_{kk} = 0$ for all k, $C_{kl} = 1$ if counties k and l share a common boundary, and $C_{kl} = 0$ otherwise. As mentioned previously, $\kappa_z \in (\lambda_{min}^{-1}, \lambda_{max}^{-1})$ to ensure that $\boldsymbol{R}_z$ is nonsingular and it turns out that for the Missouri neighborhood structure $\lambda_{min}^{-1} = -0.3457$ and $\lambda_{max}^{-1} = 0.1756$. The same general spatial structure assumed for $\boldsymbol{z}$ is also assumed for $\boldsymbol{x}$. Therefore, a multivariate normal distribution whose covariance matrix is $\sigma_x^2(\boldsymbol{I}_n - \kappa_x \boldsymbol{C})^{-1}$ is used to generate realizations of $\boldsymbol{x}$.

In order to compute (9) and (11) it remains to specify values for σ_z^2, σ_x^2, σ^2, ρ, κ_z, and κ_x. We fixed $\sigma_z^2 = 0.03$ wich corresponds to the estimated value in He and Sun (2000). To include the cases that $\eta < 1$, $\eta = 1$, and $\eta > 1$ we set $\sigma^2 \in \{0.003, 0.03, 0.3\}$. To consider similar cases for $b = \rho\dfrac{\sigma_z}{\sigma_x}$ we set $\sigma_x^2 \in \{0.003, 0.03, 0.3\}$ and $\rho \in \{0.0, 0.5, 0.9\}$ with $\rho = 0$ indicating independence between $\boldsymbol{z}$ and $\boldsymbol{x}$. Finally, we consider $\kappa_z \in \{-0.3, 0, 0.17\}$ and $\kappa_x \in \{-0.3, 0, 0.17\}$.

We generated 1000 $\boldsymbol{x}$ vectors and for each computed (9) and (11) and then averaged. Therefore, the values found in Figures 1 and 2 correspond to estimates of $E_x[\text{MSE}(\hat{\beta}_1^G \mid \boldsymbol{x}, \rho)]$ and $E_x[\text{bias}(\hat{\beta}_1^G \mid \boldsymbol{x}, \rho)]$. For a few variance component factor combinations the horizontal scale of Figures 1 and 2 masked differences between bias and MSE of $\hat{\beta}_1^G$. Therefore, we provide tables in the online supplementary material the contain values used in the figures.

Notice that in Figure 1 when $\kappa_z = \kappa_x$, then $\text{bias}(\hat{\beta}_1^G \mid \boldsymbol{x}, \rho) = \rho\dfrac{\sigma_z}{\sigma_x}$ which corroborates (10). It is curious that for a fixed κ_z, the κ_x that produces the largest bias and MSE of $\hat{\beta}_1^G$ is that when there is no spatial structure in $\boldsymbol{x}$ (i.e., $\kappa_x = 0$). However the same cannot be said for a fixed κ_x. Thus, it appears that $\boldsymbol{z}$ and $\boldsymbol{x}$ do not influence the MSE and bias equally. In line with Paciorek (2010) and Hanks et al. (2015) it appears that spatial scale (or effective range) is an important factor as moving across κ_x and/or κ_z seems to produce large changes

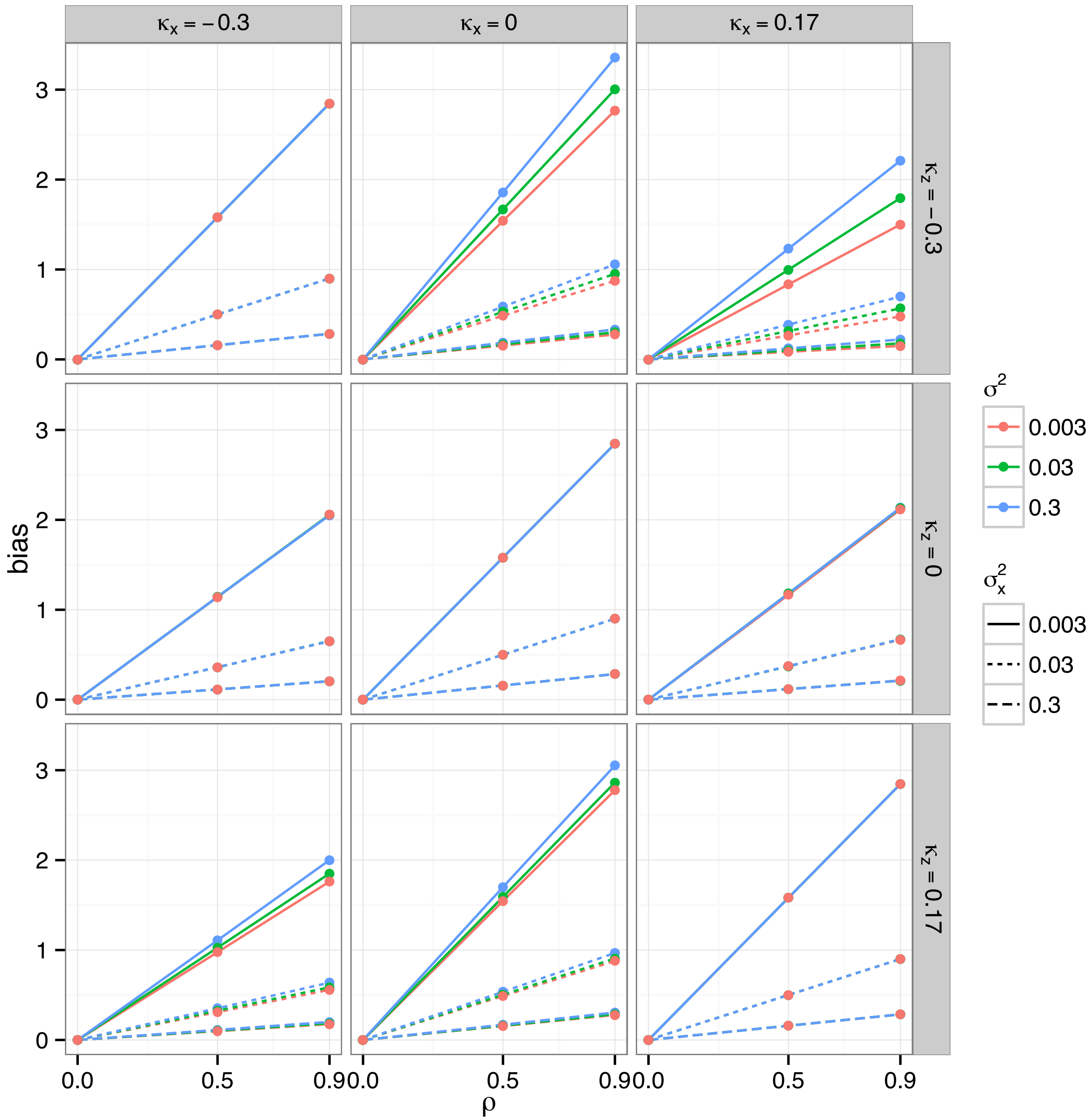

Figure 1: Numerical results for $E_x[\text{bias}(\hat{\beta}_1^G \mid x, \rho)]$ using the spatial structure available from the Missouri counties where neighborhoods are defined by counties that share a boundary. Results are averages over 1000 replicates of $\boldsymbol{x} \sim N_{114}(\mathbf{0}, \sigma_x^2(\boldsymbol{I}_n - \kappa_x \boldsymbol{C})^{-1})$

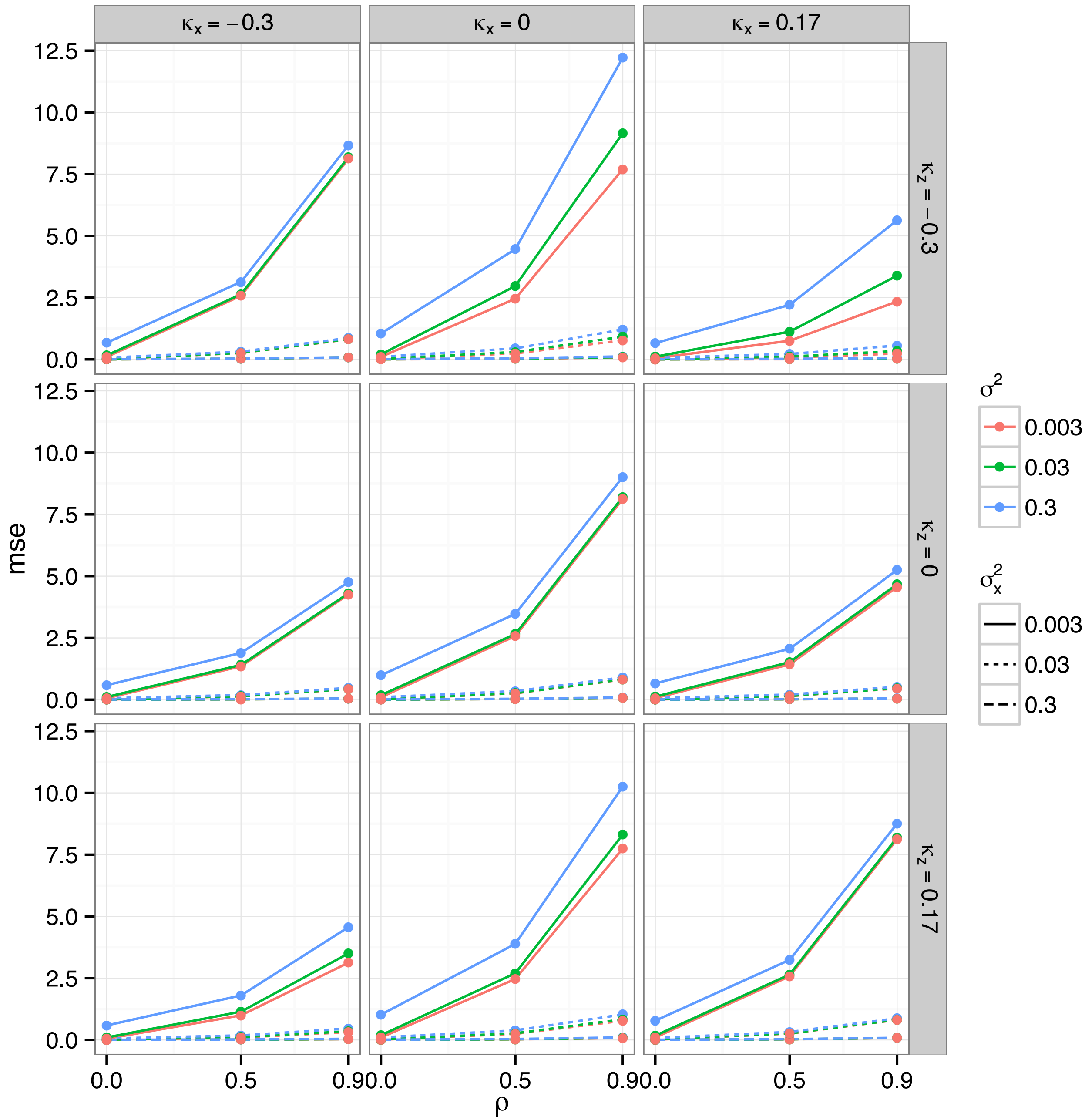

Figure 2: Numerical results for $E_x[\text{MSE}(\hat{\beta}_1^G \mid x, \rho)]$ using the spatial structure available from the Missouri counties where neighborhoods are defined by counties that share a boundary. Results are averages over 1000 replicas of $\boldsymbol{x} \sim N_{114}(\boldsymbol{0}, \sigma_x^2(\boldsymbol{I}_n - \kappa_x \boldsymbol{C})^{-1})$

in MSE and bias (though σ_x^2 appears to produce the largest effect). Finally, as expected, the bias and MSE increase as ρ increases regardless of values of other parameters.

To see how different neighborhood structures might influence bias and MSE of $\hat{\beta}_1^G$, we also considered a "queen's move" neighborhood structure on a regular grid of spatial locations and results are similar to those just detailed. See the online supplementary material for more details.

To further explore spatial scale's influence on the MSE and bias of $\hat{\beta}_1^G$, we now consider a geostatistical example since an (effective) range is explicitly parameterized in most spatial correlation functions. The numerical experiment is carried out by considering a regular grid of 25 spatial units that are located on the unit square and using pair-wise distances to construct $\boldsymbol{R}_z$ and $\boldsymbol{R}_x$. An exponential correlation function is employed which has the following form

$$\text{Corr}(\boldsymbol{s}_i, \boldsymbol{s}_j; \theta) = \exp\{-\theta ||\boldsymbol{s}_i - \boldsymbol{s}_j||\},$$

where θ is a scale parameter. The effective range for this correlation function is $d_0 \approx 3/\theta$ where d_0 denotes the distance at which the correlation drops to 0.05 (see Banerjee et al. 2014 pg. 27). We consider a sequence of values for θ so that the effective range associated with $\boldsymbol{z}$ and $\boldsymbol{x}$ extends from 0.1 to 1.99 and is incremented by 0.03. We fix $\rho \in \{0.5, 0.9\}$ and the remaining variance parameters to $\sigma^2 = \sigma_x^2 = \sigma_z^2 = 1$. For each combination of θ_z and θ_x we generated 500 $\boldsymbol{x}$ vectors and averaged (9) and (11). The results are presented in Figure 3. To highlight how $\boldsymbol{x}$ and $\boldsymbol{y}$'s effective range influences bias and MSE differently as ρ increases, the legends in the figure are not on the same scale.

From Figure 3 first notice that increasing ρ increases both bias and MSE of $\hat{\beta}_1^G$. However, the influence that $\boldsymbol{x}$ and $\boldsymbol{y}$'s effective range have on bias does not change when ρ increases while that for MSE changes drastically. Further, it appears that increasing the effective

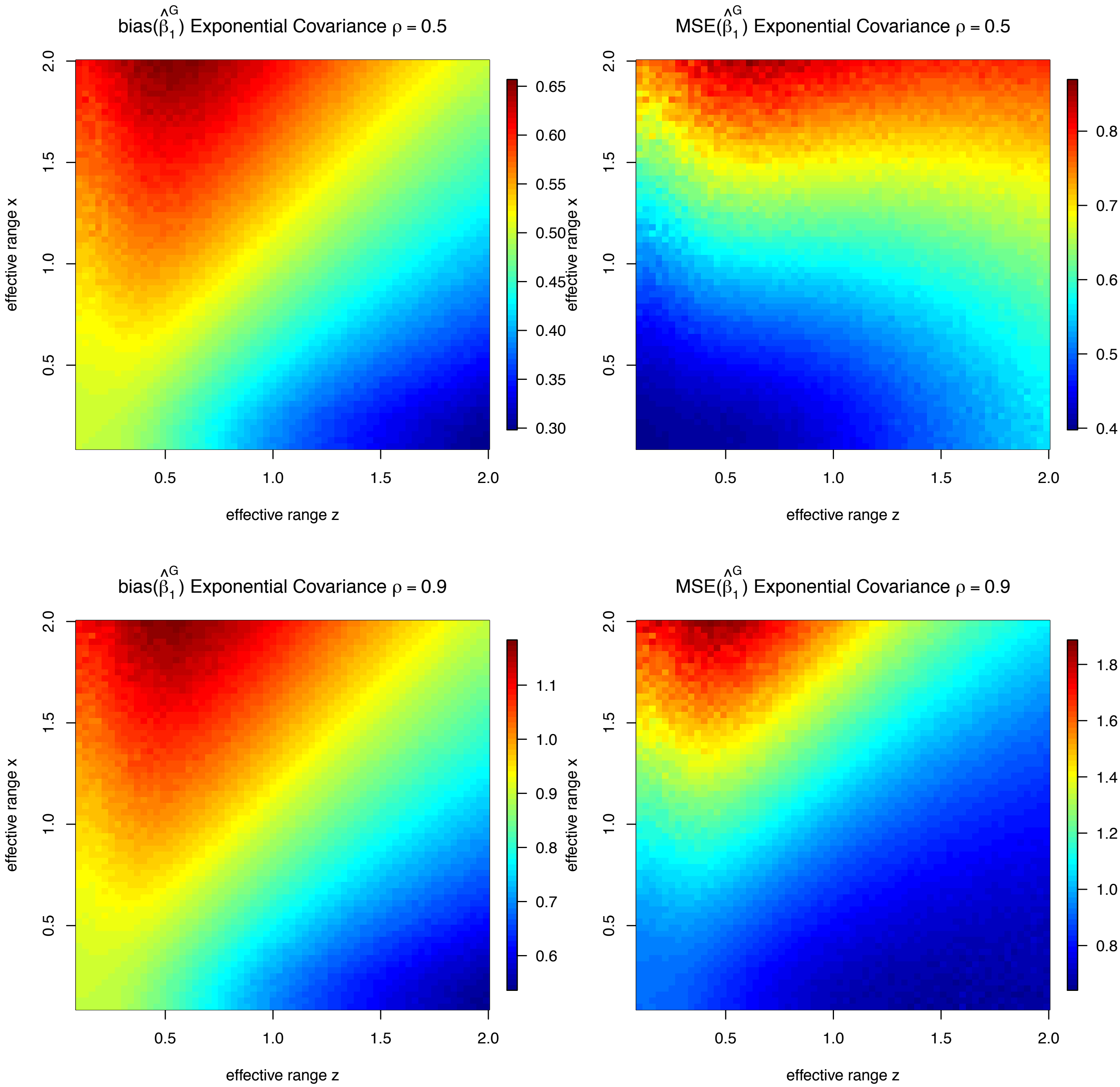

Figure 3: Bias and MSE values associated with $\hat{\beta}_1^G$. Bias and MSE were evaluated for a range of spatial scale values for both $\boldsymbol{x}$ and $\boldsymbol{z}$ using an exponential correlation function. Additionally, $\rho \in \{0.5, 0.9\}$ while all other variance components were fixed at $\sigma^2 = \sigma_x^2 = \sigma_z^2 = 1$.

range of $\boldsymbol{x}$ while holding the effective range of $\boldsymbol{z}$ constant produces an increase in the bias and MSE of $\hat{\beta}_1^G$ for both $\rho = 0.5$ and $\rho = 0.9$. However, holding $\boldsymbol{x}$'s effective range constant while increasing $\boldsymbol{z}$'s results in an overall decreased bias. As pointed out by a reviewer, a possible explanation for this might be that the variation in $\boldsymbol{z}$ is more efficiently accounted for through the regression on $\boldsymbol{x}$. The association of effective range in $\boldsymbol{z}$ and MSE for a fixed $\boldsymbol{x}$ appears to be nonlinear as it seems to increase and then decrease. We feel that these findings corroborate the results from the first simulation study of Hanks et al. (2015). As effective range of $\boldsymbol{x}$ increases it seems that bias and variance of $\hat{\beta}_1$ have similar contributions to the MSE. However as the effective range of $\boldsymbol{z}$ increases it seems as if the bias increases more than the variance which would result in higher Type S error rates. Finally, as expected, increasing ρ produces more bias and a higher MSE.

3 Prediction in the Presence of Spatial Confounding

It is common in spatial statistics to formulate models like that in (1) with the goal of making predictions. Predictions can be made at existing locations and/or at locations where a response has yet to be measured. A popular random effect predictor associated with mixed models is the best linear unbiased predictor (BLUP)(see Robinson 1991). The BLUP is best in the sense that no other linear unbiased predictor has smaller variance and as a result minimizes squared error loss when all covariance parameters are known. In a spatial setting the BLUP corresponds to the kriging predictor under certain conditions (Christensen 2011). The statistical properties of this predictor under typical mixed model assumptions have been well studied (Cressie 1993). From a Bayesian perspective, predictions are derived from the so called posterior predictive distribution which is typically denoted by $p(\boldsymbol{y}_0 \mid \boldsymbol{y})$ where $\boldsymbol{y}_0$ denotes a new observation. Handcock and Stein (1993) showed that if all covariance parameters are known and a non-informative prior is used for the mean parameter (i.e.

$\pi(\boldsymbol{\beta}) \propto 1$), the Bayes predictor ($\int \boldsymbol{y}_0 p(\boldsymbol{y}_0 \mid \boldsymbol{y}) d\boldsymbol{y}_0$) is the same as the kriging predictor. Therefore, the BLUP/kriging predictor is ubiquitous. Cressie (1993) studied the statistical properties of BLUP/kriging predictors under the assumption that covariate and error are independent. We now begin to explore the impact that spatial confounding has on the variance, bias, and mean squared prediction error (MSPE) of the BLUP/kriging predictors. It will be seen that under certain conditions these predictors can in fact perform better (in terms of MSPE) in the presence of spatial confounding.

3.1 Framework for the Kriging Predictor

We take on the same approach with prediction as that under which covariate estimation was considered. Namely, we explore the statistical properties of the kriging predictor of $\boldsymbol{y}$ under data model (1) when $\boldsymbol{x}$ and $\boldsymbol{z}$ are jointly distributed as in (5). To that end, the kriging predictor of $\boldsymbol{y}$ can be expressed by

$$\hat{\boldsymbol{y}} = \boldsymbol{X}\hat{\boldsymbol{\beta}}^G + \hat{\boldsymbol{z}}, \tag{14}$$

where $\hat{\boldsymbol{z}}$ is the BLUP of $\boldsymbol{z}$ whose specific form is

$$\hat{\boldsymbol{z}} = \frac{\sigma_z^2}{\sigma^2}\boldsymbol{R}_z\boldsymbol{\Sigma}^{-1}(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}^G) = \eta\boldsymbol{R}_z\boldsymbol{\Sigma}^{-1}(\boldsymbol{I}_n - \boldsymbol{X}\boldsymbol{A})\boldsymbol{y}. \tag{15}$$

(More details can be found in Christensen 2011 or Schabenberger and Gotway 2005.) Now substituting (15) into (14) we get

$$\hat{\boldsymbol{y}} = \boldsymbol{X}\hat{\boldsymbol{\beta}}^G + \hat{\boldsymbol{z}} = \boldsymbol{X}\boldsymbol{A}\boldsymbol{y} + \eta\boldsymbol{R}_z\boldsymbol{\Sigma}^{-1}(\boldsymbol{I}_n - \boldsymbol{X}\boldsymbol{A})\boldsymbol{y}. \tag{16}$$

As discussed in Section 2 marginalizing (16) over $\boldsymbol{z}$ would not be appropriate in the case that $\boldsymbol{x}$ and $\boldsymbol{z}$ are correlated. Instead $(\boldsymbol{z} \mid \boldsymbol{x})$ should be used. The sampling distribution of $\hat{\boldsymbol{y}}$

give $\boldsymbol{x}$ and ρ upon marginalizing over (16) with $(\boldsymbol{z} \mid \boldsymbol{x})$ is provided in the next proposition.

Proposition 3. *If $\boldsymbol{x}$ and $\boldsymbol{z}$ are jointly distributed as in* (5) *with all variance components known, then the sampling distribution of $(\hat{\boldsymbol{y}} \mid \boldsymbol{\beta}, \rho, \sigma_z^2, \sigma_x^2, \sigma^2, \boldsymbol{\theta}_x, \boldsymbol{\theta}_z, \boldsymbol{x})$ is normal with mean*

$$E(\hat{\boldsymbol{y}} \mid \boldsymbol{x}, \rho) = \boldsymbol{X\beta} + \rho\frac{\sigma_z}{\sigma_x}(\boldsymbol{\Sigma}^{-1}\boldsymbol{X}\boldsymbol{A} + \eta\boldsymbol{R}_z\boldsymbol{\Sigma}^{-1})\boldsymbol{R}_z^{1/2}\boldsymbol{R}_x^{-1/2}\boldsymbol{x} \tag{17}$$

and variance

$$\mathrm{Var}(\hat{\boldsymbol{y}} \mid \boldsymbol{x}, \rho) = \sigma^2(\boldsymbol{\Sigma}^{-1}\boldsymbol{X}\boldsymbol{A} + \eta\boldsymbol{R}_z\boldsymbol{\Sigma}^{-1})\boldsymbol{\Sigma}_\rho(\boldsymbol{\Sigma}^{-1}\boldsymbol{X}\boldsymbol{A} + \eta\boldsymbol{R}_z\boldsymbol{\Sigma}^{-1})'. \tag{18}$$

Recall that $\boldsymbol{\Sigma}_\rho = \mathrm{Var}(\boldsymbol{y} \mid \boldsymbol{x}) = \boldsymbol{I}_n + \eta(1-\rho^2)\boldsymbol{R}_z = \boldsymbol{\Sigma} - \eta\rho^2\boldsymbol{R}_z \leq \boldsymbol{\Sigma}$, for all $|\rho| \leq 1$.

Proof. The proof follows arguments similar to those found in Proposition 1 which are provided in the Appendix. □

Since $\boldsymbol{y}$ and $\hat{\boldsymbol{y}}$ are both random vectors the bias and MSPE of $\hat{\boldsymbol{y}}$ are more complicated to compute than that of $\hat{\boldsymbol{\beta}}^G$. In particular $\mathrm{bias}(\hat{\boldsymbol{y}} \mid \boldsymbol{x}, \rho) \neq E(\hat{\boldsymbol{y}} \mid \boldsymbol{x}, \rho) - \boldsymbol{y}$ and $\mathrm{MSPE}(\hat{\boldsymbol{y}} \mid \boldsymbol{x}, \rho) \neq \mathrm{tr}\{\mathrm{Cov}(\hat{\boldsymbol{y}} \mid \boldsymbol{x}, \rho)\} + \mathrm{bias}(\hat{\boldsymbol{y}} \mid \boldsymbol{x}, \rho)'\mathrm{bias}(\hat{\boldsymbol{y}} \mid \boldsymbol{x}, \rho)$. Therefore we consider MSPE and bias of $\hat{\boldsymbol{y}}$ separately with the former being derived from first principles via $\mathrm{MSPE}(\hat{\boldsymbol{y}} \mid \boldsymbol{x}, \rho) = E[(\hat{\boldsymbol{y}} - \boldsymbol{y})'(\hat{\boldsymbol{y}} - \boldsymbol{y}) \mid \boldsymbol{x}, \rho]$ and the latter with $\mathrm{bias}(\hat{\boldsymbol{y}} \mid \boldsymbol{x}, \rho) = E(\hat{\boldsymbol{y}} \mid \boldsymbol{x}, \rho) - E(\boldsymbol{y} \mid \boldsymbol{x}, \rho)$. The results are provided in the next proposition.

Proposition 4. *Under the assumptions of Proposition 3 and letting $\boldsymbol{Q} = \boldsymbol{\Sigma}^{-1}(\boldsymbol{X}\boldsymbol{A} - \boldsymbol{I}_n)$ and $\boldsymbol{\mu} = \boldsymbol{X\beta} + \rho\frac{\sigma_z}{\sigma_x}\boldsymbol{R}_z^{1/2}\boldsymbol{R}_x^{-1/2}\boldsymbol{x}$, the* bias *and* MSPE *of $\hat{\boldsymbol{y}}$ are*

$$\mathrm{bias}(\hat{\boldsymbol{y}} \mid \boldsymbol{x}, \rho) = \rho\frac{\sigma_z}{\sigma_x}\boldsymbol{Q}\boldsymbol{R}_z^{1/2}\boldsymbol{R}_x^{-1/2}\boldsymbol{x} \tag{19}$$

and

$$\text{MSPE}(\hat{\boldsymbol{y}} \mid \boldsymbol{x}, \rho) = \boldsymbol{\mu}'\boldsymbol{Q}'\boldsymbol{Q}\boldsymbol{\mu} + \sigma^2 \text{tr}\{(\boldsymbol{X}\boldsymbol{A} - \boldsymbol{I}_n)'\boldsymbol{\Sigma}^{-2}(\boldsymbol{X}\boldsymbol{A} - \boldsymbol{I}_n)\boldsymbol{\Sigma}_\rho\}. \tag{20}$$

Notice that the first term on the right hand side of (20) is $[\text{bias}(\hat{\boldsymbol{y}} \mid \boldsymbol{x}, \rho)]'[\text{bias}(\hat{\boldsymbol{y}} \mid \boldsymbol{x}, \rho)]$.

Proof. See the Appendix. □

A consequence of Proposition 4 is that if $\boldsymbol{z}$ and $\boldsymbol{x}$ are correlated ($\rho \neq 0$), $\hat{\boldsymbol{y}}$ is a biased predictor of $\boldsymbol{y}$. However, we provide two sufficient conditions that make the bias disappear. First, if $\boldsymbol{x}$ and $\boldsymbol{z}$ have the same spatial structure (see next Corollary) and second, if *both* $\boldsymbol{x}$ and $\boldsymbol{z}$ have no spatial structure (see online supplementary material). More details are provided in subsequent paragraphs, but since $\text{bias}(\hat{\boldsymbol{y}} \mid \boldsymbol{x}, \rho)$ and $\text{MSPE}(\hat{\boldsymbol{y}} \mid \boldsymbol{x}, \rho)$ are complex functions of η and $\rho\frac{\sigma_z}{\sigma_x}$ numerical exploration will be required to build intuition regarding their influence. Before proceeding, we highlight (as mentioned above) that the bias and MSPE of $\hat{\boldsymbol{y}}$ can be simplified dramatically if $\boldsymbol{R}_z = \boldsymbol{R}_x$ is assumed. The result is provided in the following corollary.

Corollary 1. *Under the assumptions of Proposition 4 and* $\boldsymbol{R}_z = \boldsymbol{R}_x$*, then the* bias *and* MSPE *of* $\hat{\boldsymbol{y}}$ *give* $\boldsymbol{x}$ *and* ρ *are*

$$\text{bias}(\hat{\boldsymbol{y}} \mid \boldsymbol{x}, \rho) = 0 \tag{21}$$

and

$$\text{MSPE}(\hat{\boldsymbol{y}} \mid \boldsymbol{x}, \rho) = \sigma^2 \text{tr}\{(\boldsymbol{X}'\boldsymbol{A}' - \boldsymbol{I}_n)\boldsymbol{\Sigma}^{-2}(\boldsymbol{X}\boldsymbol{A} - \boldsymbol{I}_n)\boldsymbol{\Sigma}_\rho\}. \tag{22}$$

Proof. See the Appendix □

From Corollary (1) we see that when spatial structure between $\boldsymbol{x}$ and $\boldsymbol{z}$ is identical, then the kriging predictor is unbiased regardless of whether spatial confounding is present or not! Looking at the form of the bias of $\hat{\boldsymbol{y}}$ it seems that when spatial structure (scale) in covariate and random effect are the same they tend to cancel each other out resulting in non-spatial collinearity under which the linear predictor is unbiased. Further, it can be shown (see the Appendix) that bias$(\hat{\boldsymbol{y}} \mid \boldsymbol{x}, \rho) = 0$ for all $\boldsymbol{x}$ if and only if $\boldsymbol{X}\boldsymbol{A} = \boldsymbol{I}_n$. Notice that if $\boldsymbol{X}\boldsymbol{A} = \boldsymbol{I}_n$ then (16) becomes $\hat{\boldsymbol{y}} = \boldsymbol{y}$ which implies that the kriging predictor under model (5) becomes an exact interpolator. In a regular spatial setting, the noiseless kriging predictor (kriging without a nugget) is also an exact interpolator (see Banerjee et al. 2014). Therefore, it is reasonable to believe that when $\boldsymbol{x}$ and $\boldsymbol{y}$ have the same spatial structure, employing the joint model (5) produces the same predictor as that derived in a typical spatial prediction setting. It therefore appears that spatial scale is crucial to understanding spatial confounding's influence on prediction. This same observation was made by Paciorek (2010) regarding spatial confounding's influence on coefficient estimation. Finally, notice that comparing MSPE when $\rho = 0$ and $\rho \neq 0$ is simply a comparison of $\boldsymbol{\Sigma}_\rho$ and $\boldsymbol{\Sigma}$

$$\text{MSPE}(\hat{\boldsymbol{y}} \mid \boldsymbol{x}, \rho) - \text{MSPE}(\hat{\boldsymbol{y}} \mid \boldsymbol{x}, \rho = 0) = \boldsymbol{\mu}'\boldsymbol{Q}'\boldsymbol{Q}\boldsymbol{\mu} + \sigma^2 \text{tr}\{\boldsymbol{Q}'\boldsymbol{Q}[\boldsymbol{\Sigma}_\rho - \boldsymbol{\Sigma}]\}.$$

3.2 Numerical Results Associated with $\hat{\boldsymbol{y}}$

As in the previous sections, to better understand how bias$(\hat{\boldsymbol{y}} \mid \boldsymbol{x}, \rho)$ and MSPE$(\hat{\boldsymbol{y}} \mid \boldsymbol{x}, \rho)$ depend on the parameters found in the expressions of Proposition 4 it is necessary to conduct a numerical study. We take on the same Missouri county neighborhood structure as in Section 2.2 and collect 1000 draws of $\boldsymbol{x}$. Because the bias of $\hat{\boldsymbol{y}}$ is a 114-dimensional vector, there are a number of methods that can be used to arrive at a single numerical summary. Among these is the absolute bias averaged over 114 counties which we employ. Therefore the values provided in Figures 4 and 5 correspond to estimates of $\frac{1}{114}\sum_{i=1}^{114} E_x(|\text{bias}(\hat{y}_i|\boldsymbol{x}, \rho)|)$. We highlight a few trends.

First focusing on bias, as expected when $\rho = 0$ or $\kappa_x = \kappa_z$ (which implies that $\boldsymbol{R}_x = \boldsymbol{R}_z$) there is no bias demonstrating (21). Additionally, as b (or ρ) increases, then the bias also increases. Further, when the difference between κ_z and κ_x increases (or spatial scales become less similar) the bias increases. Even though the influence that σ_x^2 appears to be negligible, the tables provided in the supplemental material do provide a more fine tuned picture of its impact.

Some very interesting (and unexpected) patterns emerged regarding MSPE. A decrease in η results in an increased MSPE. For a fixed κ_z, as the difference between κ_z and κ_x increases, the MPSE also increases. It is very interesting to note that the MSPE *decreases* when the absolute value of ρ increases which is the opposite to what occurred for the bias. Therefore, the decrease in variance as a result of a large ρ seems to overpower the increase in squared bias. The implication of this result is that in the presence of spatial confounding the kriging predictor might actually perform better.

In the supplementary material we provide graphs of numerical results associated with bias and MSPE of $\hat{\boldsymbol{y}}$ under a queen's move neighborhood structure. The same general trends that were discussed for the Missouri county neighborhood structure appear there as well. However, in the supplementary material we also varied σ_z^2 in addition to the other variance components and it appears that as σ_z^2 decreases ρ's ability to decrease the MSPE is diminished.

Similar to how spatial scale was explored for coefficient estimation (see Section 2.2), we employ the exponential correlation function on a 25×25 regular grid to investigate spatial scale's influence on bias$(\hat{\boldsymbol{y}} \mid \boldsymbol{x}, \rho)$ and MSPE$(\hat{\boldsymbol{y}} \mid \boldsymbol{x}, \rho)$. In Figure 6 the bias associated with $\hat{\boldsymbol{y}}$ is zero when effective range in $\boldsymbol{x}$ and $\boldsymbol{z}$ is equal and increases as ρ increases. It also appears that bias increases as the spatial range in $\boldsymbol{x}$ and $\boldsymbol{y}$ become more distant. Alternatively, as ρ increases the MSPE decreases agreeing with previous numerical results. It also appears that

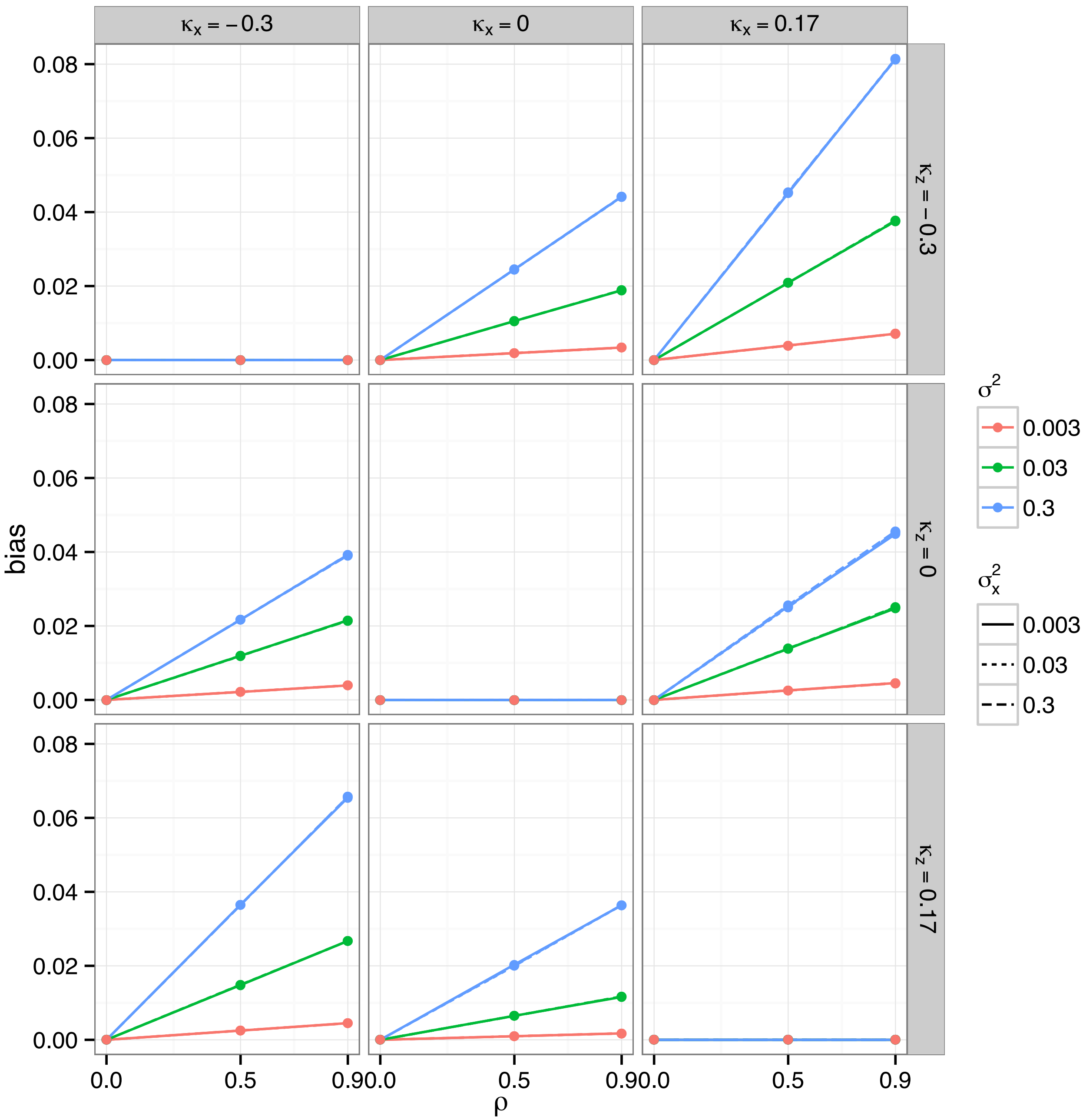

Figure 4: Numerical results for $\text{bias}(\hat{\boldsymbol{y}} \mid \boldsymbol{x}, \rho)$ using the spatial structure available from the Missouri counties where neighborhoods are defined by counties that share a boundary. Results are averages over 1000 replicas of $\boldsymbol{x} \sim N_{114}(\mathbf{0}, \sigma_x^2(\boldsymbol{I}_n - \kappa_x \boldsymbol{C})^{-1})$ and $\frac{1}{114}\sum_{i=1}^{114} E_x(|\text{bias}(\hat{y}_i|\boldsymbol{x}, \rho)|)$

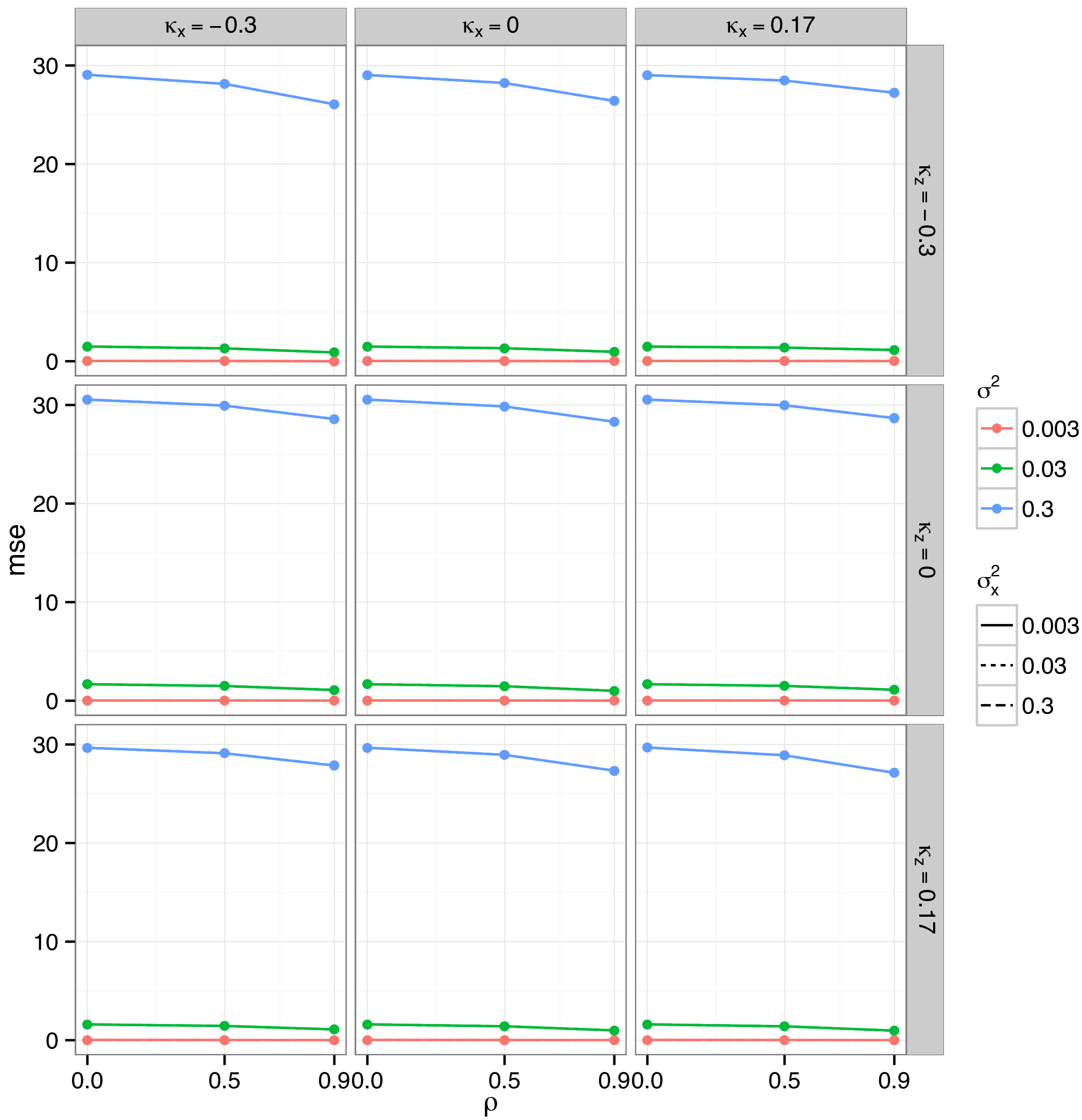

Figure 5: Numerical results for $E_x[\text{MSPE}(\hat{\boldsymbol{y}} \mid \boldsymbol{x}, \rho)]$ using the spatial structure available from the Missouri counties where neighborhoods are defined by counties that share a boundary. Results are averages over 1000 replicas of $\boldsymbol{x} \sim N_{114}(\mathbf{0}, \sigma_x^2(\boldsymbol{I}_n - \kappa_x \boldsymbol{C})^{-1})$

the effective range in $\boldsymbol{z}$ has a larger impact on the MSPE than the effective range in $\boldsymbol{x}$. This is similar to what was found in coefficient estimation.

3.3 Out of Sample Prediction of $\boldsymbol{y}$

The same procedure developed in the previous section can be used to assess spatial confounding's impact on the kriging predictor at a new location. To see this, let $\boldsymbol{s}_0$ denote a new location at which prediction is desired and let $\boldsymbol{x}_0 = (1, x_0)$ denote the covariate value measured at location $\boldsymbol{s}_0$. Further, let $\hat{y}_0$ denote the prediction for $\boldsymbol{x}_0$ at $\boldsymbol{s}_0$. Let $K(\boldsymbol{s}_0, \cdot)$ denote a valid correlation function and $\boldsymbol{r}_0 = (K(\boldsymbol{s}_0, \boldsymbol{s}_1), \ldots, K(\boldsymbol{s}_0, \boldsymbol{s}_n))$ denote the covariance between $\boldsymbol{s}_0$ and the n observed locations. Assuming the same spatial structure at the new location the kriging predictor is (see Cressie 1993)

$$\begin{aligned} \hat{y}_0 &= \boldsymbol{x}_0'\hat{\boldsymbol{\beta}} + \boldsymbol{r}_0'\boldsymbol{\Sigma}^{-1}(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}) \\ &= \boldsymbol{G}_0\boldsymbol{y}, \end{aligned} \tag{23}$$

where

$$\boldsymbol{G}_0 = \boldsymbol{x}_0'\boldsymbol{A} + \boldsymbol{r}_0'\boldsymbol{\Sigma}^{-1}(\boldsymbol{I} - \boldsymbol{X}\boldsymbol{A}). \tag{24}$$

Connecting the out of sample predictor (24) with (16) is straightforward and results would then follow in a similar manner as in Section 3.1.

4 Discussion

As pointed out in Hodges and Reich (2010), spatial confounding is pervasive in models that incorporate the idea that units near each other (in a spatial sense) produce measurements

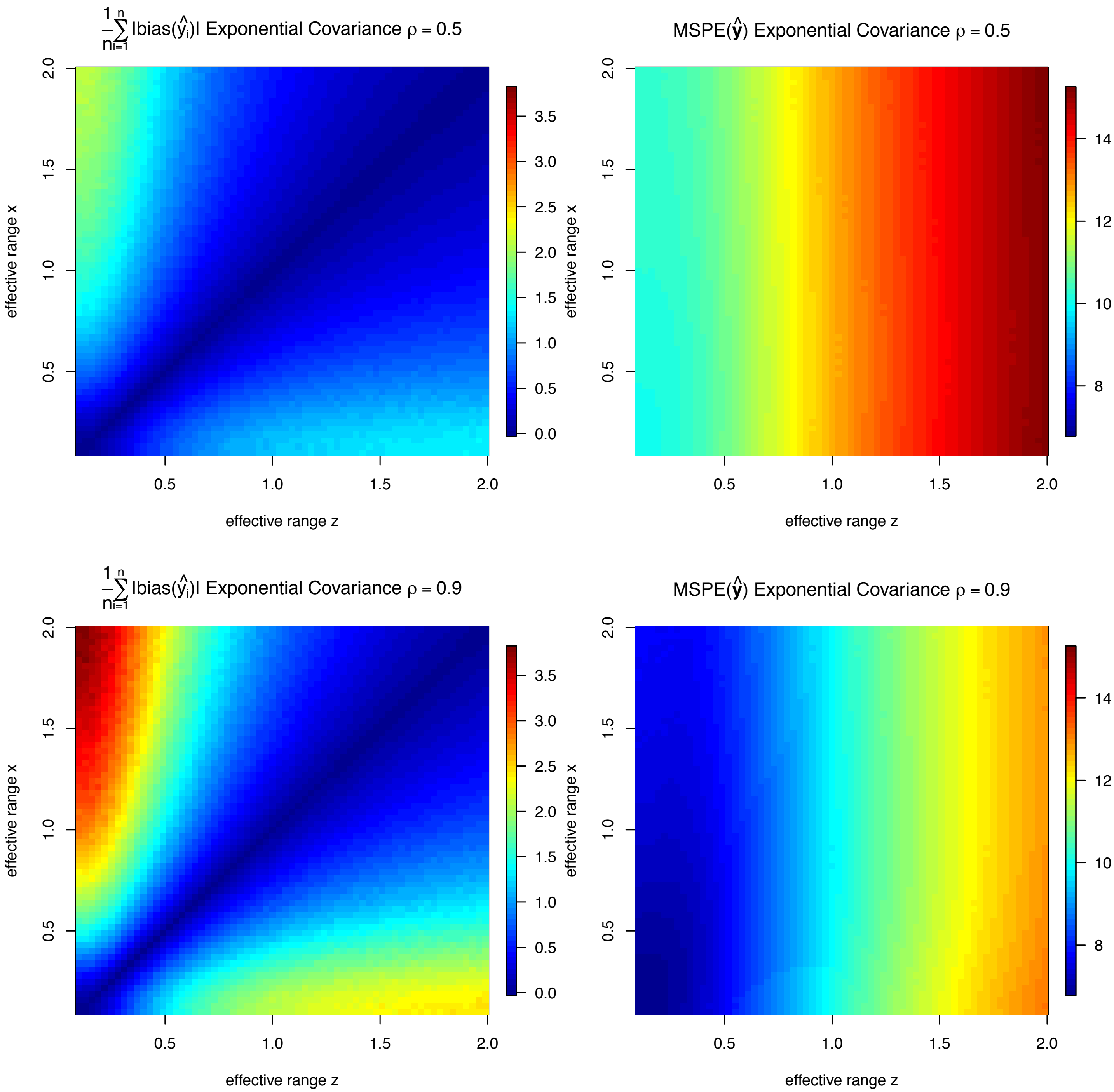

Figure 6: Bias and MSPE values associated with $\hat{\boldsymbol{y}}$. The bias and MSPE were evaluated for a range of spatial scale values for both $\boldsymbol{x}$ and $\boldsymbol{z}$ using an exponential correlation function. Additionally, $\rho \in \{0.5, 0.9\}$ while all other variance components were fixed at $\sigma^2 = \sigma_x^2 = \sigma_z^2 = 1$.

that are more similar than units that are far apart. Therefore, addressing spatial confounding is an important and necessary exercise for both applied and theoretical practitioners of spatial statistics. We've provided more insight regarding spatial confounding's impact on coefficient estimation by generalizing theory initiated in Paciorek (2010) with an accompanying numerical study. We found that the spatial scales of both response ($\boldsymbol{y}$) and covariate ($\boldsymbol{x}$) are highly influential and in some instances interact in counter intuitive ways. The more novel contributions we provide are the results regarding spatial confounding's impact on the BLUP (kriging) spatial predictors. Unlike in the *iid* setting these predictors are not unbiased and we showed that the bias depends crucially on spatial scale in $\boldsymbol{x}$ and $\boldsymbol{y}$. Further, we showed through numerical experiments that spatial confounding can actually improve prediction performance in terms of MSPE.

A natural extension to the work carried out in this article would be to explore if results hold for data that are not Gaussian. This would require studying spatial generalized linear mixed model framework (the framework under which Hughes and Haran (2013) and Hanks et al. (2015) studied spatial confounding). The theory developed in this paper depends quite heavily on the linear properties of the Gaussian model which would require new theory for data that is not Gaussian. That said, our intuition leads us to believe that the same trends would appear for non Gaussian data, but theoretical results will be hard to come by, requiring more numerical studies to learn about the influence. Finally, in practice, variance parameters will not be known and this most likely will influence bias and MSPE of the kriging predictor. What we have presented is a nice beginning to the study of spatial confounding's influence on prediction, but more research is needed to study the properties of kriging predictor when variance parameters are unknown.

Acknowledgements: We thank anonymous referees for comments that greatly improved the paper. Garritt L. Page's work was partially support by grant FONDECYT 11121131. Zhuoqiong He's research was partially supported by a grant from Missouri Department of

Conservation. Dongchu Sun's research was partially supported by the US National Science Foundation grants DMS-1007874 and SES-1260806, and by the China 111 Project (B14019).

Appendices

A Proofs

A.1 Proof of Proposition (1)

Proof. Because $(\boldsymbol{z} \mid \boldsymbol{x})$ and $\boldsymbol{e}$ are independent normal distributions, the distribution of $(\boldsymbol{y} \mid \boldsymbol{x})$ (with all variance components known) is normal with mean $\boldsymbol{X\beta} + \mathrm{E}(\boldsymbol{z} \mid \boldsymbol{x})$ and covariance $\mathrm{Var}(\boldsymbol{z} \mid \boldsymbol{x}) + \mathrm{Var}(\boldsymbol{e})$. Therefore, $\hat{\boldsymbol{\beta}}^G$ (which is a linear combination of $\boldsymbol{y}$) is normally distributed with mean

$$\begin{aligned}
\mathrm{E}(\hat{\boldsymbol{\beta}}^G \mid \boldsymbol{x}, \rho) &= (\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\mathrm{E}(\boldsymbol{y} \mid \boldsymbol{x}, \rho) \\
&= (\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\big[\boldsymbol{X\beta} + \mathrm{E}(\boldsymbol{z} \mid \boldsymbol{x}, \rho)\big] \\
&= \boldsymbol{\beta} + \rho\frac{\sigma_z}{\sigma_x}(\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{R}_z^{1/2}\boldsymbol{R}_x^{-1/2}\boldsymbol{x} \\
&= \boldsymbol{\beta} + \rho\frac{\sigma_z}{\sigma_x}\boldsymbol{A}\boldsymbol{R}_z^{1/2}\boldsymbol{R}_x^{-1/2}\boldsymbol{x},
\end{aligned}$$

and covariance matrix

$$\begin{aligned}\mathrm{Var}(\hat{\boldsymbol{\beta}}^G \mid \boldsymbol{x}, \rho) &= (\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\mathrm{Var}(\boldsymbol{y} \mid \boldsymbol{x}, \rho)\boldsymbol{\Sigma}^{-1}\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X})^{-1} \\ &= \boldsymbol{A}\big[\mathrm{Var}(\boldsymbol{z} \mid \boldsymbol{x}, \rho) + \mathrm{Var}(\boldsymbol{e} \mid \boldsymbol{x}, \rho)\big]\boldsymbol{A}' \\ &= \boldsymbol{A}\big[\sigma_z^2(1-\rho^2)\boldsymbol{R}_z + \sigma^2\boldsymbol{I}_n\big]\boldsymbol{A}' \\ &= \boldsymbol{A}(\sigma^2\boldsymbol{\Sigma} - \rho^2\sigma_z^2\boldsymbol{R}_z)\boldsymbol{A}' \\ &= \sigma^2[(\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X})^{-1} - \rho^2\eta\boldsymbol{A}\boldsymbol{R}_z\boldsymbol{A}']. \end{aligned} \tag{25}$$

□

A.2 Proof of Proposition (3)

Proof. Using similar arguments to those in Section A.1, we have

$$\begin{aligned} E(\hat{\boldsymbol{y}} \mid \boldsymbol{x}, \rho) &= [\boldsymbol{X}\boldsymbol{A} + \eta\boldsymbol{R}_z\boldsymbol{\Sigma}^{-1}(\boldsymbol{I}_n - \boldsymbol{X}\boldsymbol{A})]E(\boldsymbol{y} \mid \boldsymbol{x}, \rho) \\ &= [\boldsymbol{X}\boldsymbol{A} + \eta\boldsymbol{R}_z\boldsymbol{\Sigma}^{-1}(\boldsymbol{I}_n - \boldsymbol{X}\boldsymbol{A})](\boldsymbol{X}\boldsymbol{\beta} + \rho\frac{\sigma_z}{\sigma_x}\boldsymbol{R}_z^{1/2}\boldsymbol{R}_x^{-1/2}\boldsymbol{x}) \\ &= \boldsymbol{X}\boldsymbol{\beta} + \rho\frac{\sigma_z}{\sigma_x}[\boldsymbol{X}\boldsymbol{A} + \eta\boldsymbol{R}_z\boldsymbol{\Sigma}^{-1}(\boldsymbol{I}_n - \boldsymbol{X}\boldsymbol{A})]\boldsymbol{R}_z^{1/2}\boldsymbol{R}_x^{-1/2}\boldsymbol{x} \\ &= \boldsymbol{X}\boldsymbol{\beta} + \rho\frac{\sigma_z}{\sigma_x}[\boldsymbol{\Sigma}\boldsymbol{\Sigma}^{-1}\boldsymbol{X}\boldsymbol{A} + \eta\boldsymbol{R}_z\boldsymbol{\Sigma}^{-1} - \eta\boldsymbol{R}_z\boldsymbol{\Sigma}^{-1}\boldsymbol{X}\boldsymbol{A}]\boldsymbol{R}_z^{1/2}\boldsymbol{R}_x^{-1/2}\boldsymbol{x} \\ &= \boldsymbol{X}\boldsymbol{\beta} + \rho\frac{\sigma_z}{\sigma_x}[(\boldsymbol{\Sigma} - \eta\boldsymbol{R}_z)\boldsymbol{\Sigma}^{-1}\boldsymbol{X}\boldsymbol{A} + \eta\boldsymbol{R}_z\boldsymbol{\Sigma}^{-1}]\boldsymbol{R}_z^{1/2}\boldsymbol{R}_x^{-1/2}\boldsymbol{x} \\ &= \boldsymbol{X}\boldsymbol{\beta} + \rho\frac{\sigma_z}{\sigma_x}[\boldsymbol{\Sigma}^{-1}\boldsymbol{X}\boldsymbol{A} + \eta\boldsymbol{R}_z\boldsymbol{\Sigma}^{-1}]\boldsymbol{R}_z^{1/2}\boldsymbol{R}_x^{-1/2}\boldsymbol{x} \end{aligned} \tag{26}$$

and

$$\begin{aligned}\text{Var}(\hat{\boldsymbol{y}} \mid \boldsymbol{x}, \rho) &= [\boldsymbol{X}\boldsymbol{A} + \eta \boldsymbol{R}_z \boldsymbol{\Sigma}^{-1}(\boldsymbol{I}_n - \boldsymbol{X}\boldsymbol{A})]\text{Var}(\boldsymbol{y} \mid \boldsymbol{x}, \rho)[\boldsymbol{X}\boldsymbol{A} + \eta \boldsymbol{R}_z \boldsymbol{\Sigma}^{-1}(\boldsymbol{I}_n - \boldsymbol{X}\boldsymbol{A})]' \\ &= \sigma^2(\boldsymbol{\Sigma}^{-1}\boldsymbol{X}\boldsymbol{A} + \eta \boldsymbol{R}_z \boldsymbol{\Sigma}^{-1})\boldsymbol{\Sigma}_\rho(\boldsymbol{\Sigma}^{-1}\boldsymbol{X}\boldsymbol{A} + \eta \boldsymbol{R}_z \boldsymbol{\Sigma}^{-1})' \end{aligned} \tag{27}$$

□

A.3 Proof of Proposition (4)

Proof. Because the conditional distribution of $\boldsymbol{y}$ given $\boldsymbol{x}$ is $N(\boldsymbol{\mu}, \sigma^2\boldsymbol{\Sigma}_\rho)$ with

$$\begin{aligned}\boldsymbol{\mu} &= \boldsymbol{X}\boldsymbol{\beta} + \rho\frac{\sigma_z}{\sigma_x}\boldsymbol{R}_z^{1/2}\boldsymbol{R}_x^{-1/2}\boldsymbol{x} \\ \boldsymbol{\Sigma}_p &= \boldsymbol{I}_n + \eta(1-\rho^2)\boldsymbol{R}_z = \boldsymbol{\Sigma} - \eta\rho^2\boldsymbol{R}_z\end{aligned}$$

and

$$\hat{\boldsymbol{y}} - \boldsymbol{y} = (\boldsymbol{\Sigma}^{-1}\boldsymbol{X}\boldsymbol{A} + \eta\boldsymbol{R}_z\boldsymbol{\Sigma}^{-1} - \boldsymbol{I}_n)\boldsymbol{y} = \boldsymbol{\Sigma}^{-1}(\boldsymbol{X}\boldsymbol{A} - \boldsymbol{I}_n)\boldsymbol{y} = \boldsymbol{Q}\boldsymbol{y}$$

where $\boldsymbol{Q} = \boldsymbol{\Sigma}^{-1}(\boldsymbol{X}\boldsymbol{A} - \boldsymbol{I}_n)$ we have

$$\begin{aligned}\text{bias}(\hat{\boldsymbol{y}} \mid \boldsymbol{x}, \rho) &= E(\hat{\boldsymbol{y}} - \boldsymbol{y} \mid \boldsymbol{x}, \rho) = E(\boldsymbol{Q}\boldsymbol{y} \mid \boldsymbol{x}, \rho) = \boldsymbol{Q}E(\boldsymbol{y} \mid \boldsymbol{x}, \rho) \\ &= \boldsymbol{\Sigma}^{-1}(\boldsymbol{X}\boldsymbol{A} - \boldsymbol{I}_n)(\boldsymbol{X}\boldsymbol{\beta} + \rho\frac{\sigma_z}{\sigma_x}\boldsymbol{R}_z^{1/2}\boldsymbol{R}_x^{-1/2}\boldsymbol{x}) \\ &= \rho\frac{\sigma_z}{\sigma_x}\boldsymbol{\Sigma}^{-1}(\boldsymbol{X}\boldsymbol{A} - \boldsymbol{I}_n)\boldsymbol{R}_z^{1/2}\boldsymbol{R}_x^{-1/2}\boldsymbol{x} \\ &= \rho\frac{\sigma_z}{\sigma_x}\boldsymbol{Q}\boldsymbol{R}_z^{1/2}\boldsymbol{R}_x^{-1/2}\boldsymbol{x}\end{aligned}$$

and

$$\begin{aligned}\text{MSPE}(\hat{\boldsymbol{y}} \mid \boldsymbol{x}, \rho) &= E[(\hat{\boldsymbol{y}} - \boldsymbol{y})'(\hat{\boldsymbol{y}} - \boldsymbol{y}) \mid \boldsymbol{x}, \rho] = E[(\boldsymbol{y}'\boldsymbol{Q}'\boldsymbol{Q}\boldsymbol{y}) \mid \boldsymbol{x}, \rho] \\ &= \boldsymbol{\mu}'\boldsymbol{Q}'\boldsymbol{Q}\boldsymbol{\mu} + \sigma^2 \text{tr}\{\boldsymbol{Q}'\boldsymbol{Q}\boldsymbol{\Sigma}_\rho\} \\ &= \boldsymbol{\mu}'\boldsymbol{Q}'\boldsymbol{Q}\boldsymbol{\mu} + \sigma^2 \text{tr}\{(\boldsymbol{X}'\boldsymbol{A}' - \boldsymbol{I}_n)\boldsymbol{\Sigma}^{-2}(\boldsymbol{X}\boldsymbol{A} - \boldsymbol{I}_n)[\boldsymbol{\Sigma} - \eta\rho^2\boldsymbol{R}_z]\}.\end{aligned}$$

□

A.4 Proof of Corollary (1)

Proof.

$$\text{bias}(\hat{\boldsymbol{y}} \mid \boldsymbol{x}, \rho) = \rho\frac{\sigma_z}{\sigma_x}\boldsymbol{Q}\boldsymbol{x} = \rho\frac{\sigma_z}{\sigma_x}\boldsymbol{\Sigma}^{-1}(\boldsymbol{X}\boldsymbol{A} - \boldsymbol{I}_n)\boldsymbol{x} = \rho\frac{\sigma_z}{\sigma_x}\boldsymbol{\Sigma}^{-1}(\boldsymbol{x} - \boldsymbol{x}) = \boldsymbol{0}.$$

To show the MSPE result notice that when $\boldsymbol{R}_z = \boldsymbol{R}_x$, then $\boldsymbol{\mu} = \boldsymbol{X}\boldsymbol{\beta} + \rho\frac{\sigma_z}{\sigma_x}\boldsymbol{x}$ and thus $\boldsymbol{Q}\boldsymbol{\mu} = \boldsymbol{\Sigma}^{-1}[\boldsymbol{X}\boldsymbol{A} - \boldsymbol{I}_n](\boldsymbol{X}\boldsymbol{\beta} + \rho\frac{\sigma_z}{\sigma_x}\boldsymbol{x}) = \boldsymbol{\Sigma}^{-1}[\boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{X}\boldsymbol{\beta} + \rho\frac{\sigma_z}{\sigma_x}\boldsymbol{x} - \rho\frac{\sigma_z}{\sigma_x}\boldsymbol{x}) = \boldsymbol{0}$. Therefore,

$$\begin{aligned}\text{MSPE}(\hat{\boldsymbol{y}} \mid \boldsymbol{x}, \rho) &= \boldsymbol{\mu}'\boldsymbol{Q}'\boldsymbol{Q}\boldsymbol{\mu} + \sigma^2 \text{tr}\{(\boldsymbol{X}'\boldsymbol{A}' - \boldsymbol{I}_n)\boldsymbol{\Sigma}^{-2}(\boldsymbol{X}\boldsymbol{A} - \boldsymbol{I}_n) \\ &= \sigma^2 \text{tr}\{(\boldsymbol{X}'\boldsymbol{A}' - \boldsymbol{I}_n)\boldsymbol{\Sigma}^{-2}(\boldsymbol{X}\boldsymbol{A} - \boldsymbol{I}_n)[\boldsymbol{\Sigma} - \eta\rho^2\boldsymbol{R}_z]\}\end{aligned}$$

□

A.5 Proof of $\text{bias}(\hat{\boldsymbol{y}} \mid \boldsymbol{x}, \rho) = 0 \ \forall \ \boldsymbol{x} \iff \boldsymbol{X}\boldsymbol{A} = \boldsymbol{I}_n$

Proof.

$$
\begin{aligned}
&\text{bias}(\hat{\boldsymbol{y}} \mid \boldsymbol{x}, \rho) = 0 \ \forall \ \boldsymbol{x} \\
&\Rightarrow \rho \frac{\sigma_z}{\sigma_x} \boldsymbol{Q}\boldsymbol{R}_z^{1/2}\boldsymbol{R}_x^{-1/2}\boldsymbol{x} = 0 \ \forall \ \boldsymbol{x} \\
&\Rightarrow \boldsymbol{\Sigma}^{-1}(\boldsymbol{X}\boldsymbol{A} - \boldsymbol{I}_n)\boldsymbol{R}_z^{1/2}\boldsymbol{R}_x^{-1/2}\boldsymbol{x} = 0 \ \forall \ \boldsymbol{x} \\
&\Rightarrow \boldsymbol{X}\boldsymbol{A}\boldsymbol{R}_z^{1/2}\boldsymbol{R}_x^{-1/2}\boldsymbol{x} = \boldsymbol{R}_z^{1/2}\boldsymbol{R}_x^{-1/2}\boldsymbol{x} \ \forall \ \boldsymbol{x} \\
&\Rightarrow \boldsymbol{X}\boldsymbol{A}\boldsymbol{R}_z^{1/2}\boldsymbol{R}_x^{-1/2} = \boldsymbol{R}_z^{1/2}\boldsymbol{R}_x^{-1/2} \\
&\Rightarrow \boldsymbol{X}\boldsymbol{A} = \boldsymbol{I}_n
\end{aligned}
$$

Now if $\boldsymbol{X}\boldsymbol{A} = \boldsymbol{I}_n$ it is obvious from (19) that $\text{bias}(\hat{\boldsymbol{y}} \mid \boldsymbol{x}, \rho) = 0 \ \forall \ \boldsymbol{x}$. □

References

Banerjee, S., Carlin, B. P., and Gelfand, A. E. (2014), *Hierarchical Modeling and Analysis for Spatial Data*, Chapman and Hall/CRC, 2nd ed.

Besag, J. and Higdon, D. (1999), "Bayesian Analysis of Agricultural Field Experiments," *Journal of the Royal Statistical Society: Series B*, 61, 691–746.

Caragea, P. C. and Kaiser, M. S. (2009), "Autologistic Models with Interpretable Parameters," *Journal of Agricultural, Biological, and Environmental Statistics*, 14, 281–300.

Christensen, R. (2011), *Plane Answers to Complex Questions: The Theory of Linear Models*, Springer Texts in Statistics, Springer, fouth ed.

Clayton, D. and Kaldor, J. (1987), "Empirical Bayes Estimates of Age-Standardized Relative Risks for Use in Disease Mapping," *Biometrics*, 43, 671–681.

Clayton, D. G., Bernardinellli, L., and Montomoli, C. (1993), "Spatial Correlation in Ecological Analysis," *International Journal of Epidemiology*, 22, 1193–1202.

Cressie, N. A. (1993), *Statistics for Spatial Data*, Whiley, 2nd ed.

Gelman, A. and Tuerlinckx, F. (2000), "Type S Error Rates for Classical and Bayesian Single and Multiple Comparison Procedures," *Computational Statistics*, 15, 373–390.

Green, P. J. (1985), "Linear Models for Field Trials, Smoothing and Cross-Validation," *Biometrika*, 72, 527–37.

Handcock, M. S. and Stein, M. L. (1993), "A Bayesian Analysis of Kriging," *Technometrics*, 35, 403–410.

Hanks, E. M., Schleip, E. M., Hooten, M. B., and Hoeting, J. A. (2015), "Restricted Spatial Regression in Practice: Geostatistical Models, Confounding, and Robustness Under Model Misspecification," *Environmentrics*, 26, 243–256.

He, Z. and Sun, D. (2000), "Hierarchical Bayes Estimation of Hunting Success Rates with Spatial Correlations," *Biometrics*, 56, 360–267.

Hodges, J. S. and Reich, B. J. (2010), "Adding Spatially-Correlated Errors Can Mess Up the Fixed Effect You Love," *The American Statistician*, 64, 325–334.

Hughes, J. and Haran, M. (2013), "Dimension Reduction and Alleviation of Confounding for Spatial Generalized Linear Mixed Models," *Journal of the Royal Statistical Society, Series B*, 75, 139–159.

Lee, D., Rushworth, A., and Sahu, S. K. (2014), "A Bayesian Localized Conditional Autoregressive Model for Estimating the Health Effects of Air Pollution," *Biometrics*, 70, 419–429.

Paciorek, C. J. (2010), "The Importance of Scale for Spatial-Confounding Bias and Precision of Spatial Regression Estimators," *Statistical Science*, 25, 107–125.

Reich, B. J., Hodges, J. S., and Zadnik, V. (2006), "Effects of Residual Smoothing on the Posterior of the Fixed Effects in Disease-Mapping Models," *Biometrics*, 62, 1197–1206.

Robinson, G. K. (1991), "That BLUP is a Good Thing: The Estimation of Random Effects," *Statistical Science*, 6, 15–51.

Schabenberger, O. and Gotway, C. A. (2005), *Statistical Methods for Spatial Data Analysis*, Chapman & Hall/CRC.

Speckman, P. (1988), "Kernel Smoothing in Partial Linear Models," *Journal of the Royal Statistical Society. Series B*, 50, 413–436.

Wakefield, J. (2007), "Disease Mapping and Spatial Regression with Count Data," *Biostatistics*, 8, 158–183.

Woodard, R., Sun, D., He, Z., and Sheriff, S. L. (1999), "Estimating Hunting Success Rates via Bayesian Generalized Linear Models," *Journal of Agricultural, Biological, and Environmental Statistics*, 4, 456–472.