JOURNAL OF MULTIVARIATE ANALYSIS 45, 56–72 (1993)

Improved Multivariate Prediction under a General Linear Model

CAROL A. GOTWAY*

Department of Biometry, University of Nebraska

AND

NOEL CRESSIE

Department of Statistics, Iowa State University

Assuming a general linear model with known covariance matrix, several linear and nonlinear predictors are presented and their properties are discussed. In the context of simultaneous multiple prediction, a total sum of squared errors is suggested as a loss function for comparing predictors. Based on a fundamental relationship between prediction and estimation, a very general class of predictors is developed from which predictors with uniformly smaller risk than that of the classical best linear unbiased (i.e., universal kriging) predictor can be constructed.

1. INTRODUCTION

Over the past 20 years, a large amount of research has been carried out in the area of shrinkage estimation. Since the introduction of the James–Stein estimator in 1961 [15], many significant extensions of their work have proven useful. Examples include extensions to linear models with general covariance structures, the use of different types of loss functions, and the application of shrinkage estimation to families of distributions other than the Gaussian.

However, it has not been until recently, through the work of Copas [7] and Copas and Jones [8], that shrinkage techniques have been extended to the prediction of a random variable. In this paper, we carry their development of shrinkage predictors one step further by developing a broad class of predictors that is based on a general linear model with

Received September 16, 1991; revised August 17, 1992.

AMS 1980 classifications: primary 62H99; secondary 62M20.

Key words and phrases: kriging, minimax prediction, multivariate inference, simultaneous prediction.

* Formerly with Sandia National Laboratories, Albuquerque, NM.

0047-259X/93 $5.00

arbitrary covariance structure. From this class, the best linear unbiased predictor, as well as the shrinkage predictors of Copas [7], may be obtained.

Assume the following general linear model for $\mathbf{Z}$, and unobservables $\mathbf{Z}_0$:

$$\begin{aligned}\mathbf{Z} &= X\boldsymbol{\beta} + \boldsymbol{\varepsilon} \\ \mathbf{Z}_0 &= X_0\boldsymbol{\beta} + \boldsymbol{\varepsilon}_0,\end{aligned} \tag{1.1}$$

where

$\mathbf{Z}$ is the n-dimensional data vector;

$\mathbf{Z}_0$ is a k-dimensional vector of unobserved values that will be predicted from the data; write $\mathbf{Z}_0 = (Z_{0,1}, ..., Z_{0,k})'$.

X and X_0 are matrices of explanatory variables ($\text{rank}(X) = p$), the rows of which can represent a trend surface;

$\boldsymbol{\beta}$ is a $p \times 1$ vector of unknown fixed parameters; and

$\boldsymbol{\varepsilon}$ and $\boldsymbol{\varepsilon}$ are random errors with zero mean and covariance matrix

$$\text{var}((\boldsymbol{\varepsilon}', \boldsymbol{\varepsilon}_0')') = \sigma^2 \begin{bmatrix} \Sigma_{ZZ} & \Sigma_{Z0} \\ \Sigma_{0Z} & \Sigma_{00} \end{bmatrix} \equiv \sigma^2 \Sigma, \tag{1.2}$$

where Σ_{ZZ}, Σ_{00}, and Σ are known positive definite matrices and σ^2 is an unknown constant.

Based on this model, the problem is to predict the k-dimensional vector $\mathbf{Z}_0$ using some optimal function of the data $\mathbf{Z}$. Let the predictor be

$$\mathbf{p}(\mathbf{Z}) = (p_1(\mathbf{Z}), ..., p_k(\mathbf{Z}))'. \tag{1.3}$$

Following the ideas of James and Stein [15], in the realm of multivariate estimation, an appropriate measure (loss function) of overall prediction performance is

$$L(\mathbf{p}, \mathbf{Z}_0) = \sigma^{-2} \sum_{i=1}^{k} (p_i(\mathbf{Z}) - Z_{0,i})^2, \tag{1.4}$$

which may be considered as an index of accuracy for either stochastic or nonstochastic predictors. For stochastic predictors, expectations may be taken, yielding the risk function

$$\begin{aligned} r(\mathbf{p}, \mathbf{Z}_0; \boldsymbol{\beta}, \sigma^2) &= \sigma^{-2} E\left(\sum_{i=1}^{k} (p_i(\mathbf{Z}) - Z_{0,i})^2\right) \\ &= \sigma^{-2} E((\mathbf{p}(\mathbf{Z}) - \mathbf{Z}_0)' (\mathbf{p}(\mathbf{Z}) - \mathbf{Z}_0)), \end{aligned} \tag{1.5}$$

which reflects the desire to predict well at all k points *collectively*.

In the following sections, we present and compare several stochastic predictors using the measure (1.5). Section 2 gives a general class of predictors based on a fundamental relationship between prediction and estimation. Section 3 considers the risk function for these predictors, and in Section 4, predictors with smaller risk than that of the best linear unbiased predictor are constructed. Section 5 compares a few of these predictors using an example based on a spatial data configuration.

2. A General Class of Predictors

Initially, to motivate the ideas behind linear prediction, we shall assume that $\boldsymbol{\beta}$ in (1.1) is known. Consider the heterogeneous linear predictor

$$\mathbf{p}(\mathbf{Z}) = B\mathbf{Z} + \mathbf{c}, \tag{2.1}$$

where B is a $k \times n$ matrix and $\mathbf{c}$ is a $k \times 1$ vector. From (1.5), its risk is

$$\begin{aligned} &\sigma^{-2}\{(BX - X_0)\boldsymbol{\beta} + \mathbf{c})' \, ((BX - X_0)\boldsymbol{\beta} + \mathbf{c})\} \\ &\quad + \operatorname{tr}(B\Sigma_{ZZ}B' + \Sigma_{00} - 2B\Sigma_{Z0}). \end{aligned} \tag{2.2}$$

By differentiating (2.2) with respect to B and $\mathbf{c}$, and equating to zero, the optimal values for B and $\mathbf{c}$ are

$$B_{\text{opt}} = \Sigma_{0Z}\Sigma_{ZZ}^{-1}, \qquad \mathbf{c}_{\text{opt}} = X_0\boldsymbol{\beta} - \Sigma_{0Z}\Sigma_{ZZ}^{-1}X\boldsymbol{\beta}; \tag{2.3}$$

that these values minimize (2.2) is shown in Toutenburg [24, p. 140]. Hence, the best heterogeneous linear predictor is

$$\mathbf{p}_1(\mathbf{Z}) = \Sigma_{0Z}\Sigma_{ZZ}^{-1}\mathbf{Z} + (X_0 - \Sigma_{0Z}\Sigma_{ZZ}^{-1}X)\boldsymbol{\beta}. \tag{2.4}$$

This predictor $\mathbf{p}_1(\mathbf{Z})$ is well known in time series (see [25, p. 77 or [10, p. 75], and in geostatistics it is called the *simple kriging* predictor.

The best heterogeneous linear predictor (2.4) is unbiased for $\mathbf{Z}_0$, and the minimized risk is

$$r(\mathbf{p}_1, \mathbf{Z}_0; \boldsymbol{\beta}, \sigma^2) = \operatorname{tr}(\Sigma_{00} - \Sigma_{0Z}\Sigma_{ZZ}^{-1}\Sigma_{Z0}). \tag{2.5}$$

Now, consider the general class of predictors suggested by (2.4), namely,

$$\mathbf{p}(\mathbf{Z}, \hat{\boldsymbol{\beta}}) = \Sigma_{0Z}\Sigma_{ZZ}^{-1} + \mathbf{Z}(X_0 - \Sigma_{0Z}\Sigma_{ZZ}^{-1}X)\hat{\boldsymbol{\beta}}, \tag{2.6}$$

where $\hat{\boldsymbol{\beta}}$ is any estimator of $\boldsymbol{\beta}$. If $(X_0 - \Sigma_{0Z}\Sigma_{ZZ}X)$ is of full column rank, then any predictor may be written in this form, for if $\mathbf{p}(\mathbf{Z})$ is any predictor, take $\hat{\boldsymbol{\beta}}$ to be

$$\{(X_0 - \Sigma_{0Z}\Sigma_{ZZ}^{-1}X)'(X_0 - \Sigma_{0Z}\Sigma_{ZZ}^{-1}X)\}^{-1}$$
$$\times (X_0 - \Sigma_{0Z}\Sigma_{ZZ}^{-1}X)'(\mathbf{p}(\mathbf{Z}) - \Sigma_{0Z}\Sigma_{ZZ}^{-1}\mathbf{Z}).$$

Moreover, $\mathbf{p}(\mathbf{Z}, \hat{\boldsymbol{\beta}})$ inherits its first-order and second-order moment properties from the estimator $\hat{\boldsymbol{\beta}}$; e.g., if $\hat{\boldsymbol{\beta}}$ is unbiased for $\boldsymbol{\beta}$, then $\mathbf{p}(\mathbf{Z}, \hat{\boldsymbol{\beta}})$ is unbiased for $\mathbf{Z}_0$.

The best linear unbiased predictor (BLUP) [12] is a member of this class and is obtained by estimating $\boldsymbol{\beta}$ with the generalized least squares estimator of $\hat{\boldsymbol{\beta}}$, namely

$$\hat{\boldsymbol{\beta}}_{\mathrm{GLS}} = (X'\Sigma_{ZZ}^{-1}X)^{-1}X'\Sigma_{ZZ}^{-1}\mathbf{Z}. \tag{2.7}$$

Then the corresponding predictor, i.e., the BLUP is

$$\mathbf{p}_2(\mathbf{Z}) \equiv \mathbf{p}(\mathbf{Z}, \hat{\boldsymbol{\beta}}_{\mathrm{GLS}}) = \Sigma_{0Z}\Sigma_{ZZ}^{-1}\mathbf{Z} + (X_0 - \Sigma_{0Z}\Sigma_{ZZ}^{-1}X)\hat{\boldsymbol{\beta}}_{\mathrm{GLS}}, \tag{2.8}$$

with the risk given by

$$r(\mathbf{p}_2, \mathbf{Z}_0; \boldsymbol{\beta}, \sigma^2) = \mathrm{tr}\{(X'\Sigma_{ZZ}^{-1}X)^{-1}Q\},$$

where Q is given in Eq. (3.3) below.

There are other important members of the class (2.6). The preshrunk predictors of Copas [7] can be obtained by taking $\Sigma_{ZZ} = I$, $\Sigma_{0Z} = \phi$ (a $k \times n$ matrix of zeros) and by taking $\hat{\boldsymbol{\beta}}$ to be

$$\hat{\boldsymbol{\beta}}_{\mathrm{OJS}} = \left(1 - \frac{(p-2)(n-p)s^2}{(n-p+2)\hat{\boldsymbol{\beta}}'_{\mathrm{OLS}}X'X\hat{\boldsymbol{\beta}}_{\mathrm{OLS}}}\right)\hat{\boldsymbol{\beta}}_{\mathrm{OLS}}, \tag{2.9}$$

where $\hat{\boldsymbol{\beta}}_{\mathrm{OLS}} = (X'X)^{-1}X'\mathbf{Z}$ and $s^2 = (1/(n-p))(\mathbf{Z} - X\hat{\boldsymbol{\beta}}_{\mathrm{OLS}})'(\mathbf{Z} - X\hat{\boldsymbol{\beta}}_{\mathrm{OLS}})$. Upon substitution into (2.6), we obtain

$$\mathbf{p}_3(Z) \equiv \mathbf{p}(\mathbf{Z}, \hat{\boldsymbol{\beta}}_{\mathrm{OJS}}) = X_0\hat{\boldsymbol{\beta}}_{\mathrm{OJS}}. \tag{2.10}$$

The predictors developed by Copas [7] were developed for the prediction of a single random variable assuming that the X_0 at which future predictions are required is also random and not a fixed quantity, as we have assumed. To obtain a reduction in risk over the BLUP for a *single* random variable, averaging over the explanatory variables is crucial. The focus of this paper is not the prediction of a single random variable, but the prediction of *many* random variables simultaneously. Although the predictors of Copas [7] are embedded in (2.6) algebraically, his approach to shrinkage prediction is different from that given here.

In geostatistical applications where the covariance structure of model (1.1) is a function of the distance between data locations, the predictors given by (2.6) honor the data. In this case, $X_0 = X$, $\Sigma_{0Z} = \Sigma_{ZZ} = \Sigma_{00}$, so that $\mathbf{p}(\mathbf{Z}, \hat{\boldsymbol{\beta}}) = \mathbf{Z}$, and $r(\mathbf{p}, \mathbf{Z}_0; \boldsymbol{\beta}, \sigma^2) = 0$. Also, when $\Sigma_{0Z} = \phi$, $\mathbf{p}(\mathbf{Z}, \hat{\boldsymbol{\beta}}) = X_0\hat{\boldsymbol{\beta}}$, and the prediction problem reduces to one of estimation of $\boldsymbol{\beta}$.

For general Σ_{ZZ} and Σ_{0Z}, the general James–Stein type estimator of $\boldsymbol{\beta}$,

$$\hat{\boldsymbol{\beta}}_{\mathrm{GJS}} = \left(1 - \frac{a(n-p)\hat{\sigma}^2}{\hat{\boldsymbol{\beta}}'_{\mathrm{GLS}} X' \Sigma_{ZZ}^{-1} X \hat{\boldsymbol{\beta}}_{\mathrm{GLS}}}\right) \hat{\boldsymbol{\beta}}_{\mathrm{GLS}} \tag{2.11}$$

may be chosen to give

$$\mathbf{p}_4(\mathbf{Z}) \equiv \mathbf{p}(\mathbf{Z}, \hat{\boldsymbol{\beta}}_{\mathrm{GJS}}) = \Sigma_{0Z}\Sigma_{ZZ}^{-1}\mathbf{Z} + (X_0 - \Sigma_{0Z}\Sigma_{ZZ}^{-1}X)\hat{\boldsymbol{\beta}}_{\mathrm{GJS}}. \tag{2.12}$$

In (2.11), a is a constant that is chosen so that the James–Stein estimator has uniformly smaller mean-squared error than that of the generalized least squares estimator, and $\hat{\sigma}^2 = 1/(n-p)(\mathbf{Z} - X\hat{\boldsymbol{\beta}}_{\mathrm{GLS}})' \Sigma_{ZZ}^{-1}(\mathbf{Z} - X\hat{\boldsymbol{\beta}}_{\mathrm{GLS}})$. Discussion of appropriate choices for a, as well as a resentation of James–Stein estimators under the framework of minimax estimation, will be presented in subsequent sections.

Clearly, many similar types of predictors may be obtained in the manner illustrated above. Other choices of $\hat{\boldsymbol{\beta}}$ that lead to predictors different from the BLUP include ridge-regression esmtimators [14], minimax adaptive generalized ridge estimators [23], and Bayes and empirical Bayes estimators [9]. For a good general discussion on estimation in linear models, see Rao [20].

The remainder of this article focuses on the construction of predictors with uniformly smaller risk than that of the best linear unbiased (kriging) predictor. In this context, minimax estimators will be considered.

3. Decomposition of Risk

Harville [13] presents a general decomposition of prediction error that can be specialized to our situation. We present here a direct proof of the decomposition of risk since it illuminates our basic approach to improved prediction.

LEMMA 3.1. *Assume the model* (1.1). *If* $\hat{\boldsymbol{\beta}} = B\mathbf{Z}$ *for some* $p \times n$ *matrix* B, *and* $(p_1(\mathbf{Z}, \hat{\boldsymbol{\beta}}), ..., p_k(\mathbf{Z}, \hat{\boldsymbol{\beta}}))'$ *given by Eq.* (2.6) *is a predictor of* $\mathbf{Z}_0 = (Z_{0,1}, ..., Z_{0,k})'$, *then*

$$E(p_i(\mathbf{Z}, \hat{\boldsymbol{\beta}}) - Z_{0,i})^2 = E(p_i(\mathbf{Z}, \boldsymbol{\beta}) - Z_{0,i})^2 + E(p_i(\mathbf{Z}, \hat{\boldsymbol{\beta}}) - p_i(\mathbf{Z}, \boldsymbol{\beta}))^2,$$

which upon expansion is equal to

$$\sigma^2(\Sigma_{00,ii} - \Sigma_{0Z,i}\Sigma_{ZZ}^{-1}(\Sigma_{0Z,i})') + E\{(\hat{\boldsymbol\beta} - \boldsymbol\beta)'(X_{0,i} - \Sigma_{0Z,i}\Sigma_{ZZ}^{-1}X)'(X_{0,i} - \Sigma_{0Z,i}\Sigma_{ZZ}^{-1}X)(\hat{\boldsymbol\beta} - \boldsymbol\beta)\},$$

where $\Sigma_{00,ii}$ is the (i, i)th element of Σ_{00} and $X_{0,i}$ and $\Sigma_{0Z,i}$ are the ith rows of X_0 and Σ_{0Z}, respectively.

Proof.

$$E(p_i(\mathbf{Z}, \hat{\boldsymbol\beta}) - Z_{0,i})^2 = E(p_i(\mathbf{Z}, \hat{\boldsymbol\beta}) - p_i(\mathbf{Z}, \boldsymbol\beta))^2 + 2E\{(p_i(\mathbf{Z}, \hat{\boldsymbol\beta}) - p_i(\mathbf{Z}, \boldsymbol\beta))(p_i(\mathbf{Z}, \boldsymbol\beta) - Z_{0,i})\} + E(p_i(\mathbf{Z}, \boldsymbol\beta) - Z_{0,i})^2.$$

If $\hat{\boldsymbol\beta} = B\mathbf{Z}$, then $\text{cov}(\hat{\boldsymbol\beta}, p_i(\mathbf{Z}, \boldsymbol\beta) - Z_{0,i}) = \mathbf{0}$ and the cross-product term of (3.1) is zero. Thus,

$$E(p_i(\mathbf{Z}, \hat{\boldsymbol\beta}) - Z_{0,i})^2 = E(p_i(\mathbf{Z}, \hat{\boldsymbol\beta}) - p_i(\mathbf{Z}, \boldsymbol\beta))^2 + E(p_i(\mathbf{Z}, \boldsymbol\beta) - Z_{0,i})^2. \quad (3.1)$$

Substituting $p_i(\mathbf{Z}, \hat{\boldsymbol\beta}) = \Sigma_{0Z,i}\Sigma_{ZZ}^{-1}\mathbf{Z} + (X_{0,i} - \Sigma_{0Z,i}\Sigma_{ZZ}^{-1}X)\hat{\boldsymbol\beta}$, and $p_i(\mathbf{Z}, \boldsymbol\beta) - \Sigma_{0Z,i}\Sigma_{ZZ}^{-1}\mathbf{Z} + (X_{0,i} - \Sigma_{0Z,i}\Sigma_{ZZ}^{-1}X)\boldsymbol\beta$ into (3.1) and taking expectations gives the result.

LEMMA 3.2. *Assume the model* (1.1) *and assume also that*

$$E(Z_{0,i} \mid \mathbf{Z}) = \Sigma_{0Z,i}\Sigma_{ZZ}^{-1}\mathbf{Z} + (X_{0,i} - \Sigma_{0Z,i}\Sigma_{ZZ}^{-1}X)\boldsymbol\beta \equiv p_i(\mathbf{Z}, \boldsymbol\beta).$$

The Lemma 3.1 *holds for any $\hat{\boldsymbol\beta}$, linear or not.*

Proof. Let $\boldsymbol\delta(\mathbf{Z})$ be any function of the data $\mathbf{Z}$. Then

$$\text{cov}(\boldsymbol\delta(\mathbf{Z}), p_i(\mathbf{Z}, \boldsymbol\beta) - Z_{0,i}) = E[\boldsymbol\delta(\mathbf{Z})(E(Z_{0,i} \mid \mathbf{Z}) - Z_{0,i})] = E[\boldsymbol\delta(\mathbf{Z})\{E(E(Z_{0,i} \mid \mathbf{Z}) - Z_{0,i}) \mid \mathbf{Z}\}] = \mathbf{0}.$$

Thus, since $\hat{\boldsymbol\beta} = \boldsymbol\delta(\mathbf{Z})$, for some function $\boldsymbol\delta$, $\text{cov}(\hat{\boldsymbol\beta}, p_i(\mathbf{Z}, \boldsymbol\beta) - Z_{0,i}) = \mathbf{0}$, and the cross-product term in (3.1) is zero. Hence the decomposition of Lemma 3.1 holds.

Clearly, when $\mathbf{Z}$ and $Z_{0,i}$ are jointly normal, the assumption of Lemma 3.2 is satisfied. However, Lemma 3.2 also holds when $\mathbf{Y}_i = (\mathbf{Z}', Z_{0,i})'$ has an elliptically symmetric distribution with location vector $\boldsymbol\mu_i = ((X\boldsymbol\beta)', X_{0,i}\boldsymbol\beta)'$, covariance matrix Σ_i, and a probability density function of the form:

$$f(\mathbf{y}) = |\Sigma|^{-1/2}\,\psi\{(\mathbf{y}_i - \boldsymbol\mu_i)'\Sigma_i^{-1}(\mathbf{y}_i - \boldsymbol\mu_i)\}.$$

Kelker [17] gives the properties of elliptically symmetric distributions and shows that the conditional expectation of $\mathbf{Z}$ given $Z_{0,i}$ has the form of that required by Lemma 3.2. Finally, if $(\mathbf{Z}', \mathbf{Z}_0')'$ has an elliptically symmetric distribution, so too does $\mathbf{Y}_i$, $i = 1, ..., k$.

Thus, whenever $\mathbf{Z}_0$ regresses linearly on $\mathbf{Z}$, Lemma 3.2 shows that the risk function (1.5) may be decomposed into the sum of two parts:

$$r(\mathbf{p}, \mathbf{Z}_0; \boldsymbol{\beta}, \sigma^2) = \operatorname{tr}(\Sigma_{00} - \Sigma_{0Z}\Sigma_{ZZ}^{-1}\Sigma_{Z0}) + \sigma^{-2}E(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})' Q(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}), \quad (3.2)$$

where

$$Q = (X_0 - \Sigma_{0Z}\Sigma_{ZZ}^{-1}X)'\,(X_0 - \Sigma_{0Z}\Sigma_{ZZ}^{-1}X). \quad (3.3)$$

The first term of the risk (3.2) is inherent in prediction of $\mathbf{Z}_0$ with $\boldsymbol{\beta}$ known (it is equal to $r(\mathbf{p}_1, \mathbf{Z}_0; \boldsymbol{\beta}, \sigma^2)$); the second is due to the estimation of $\boldsymbol{\beta}$.

The decomposition (3.2) has important implications for the comparison of two predictors.

DEFINITION. A predictor $\mathbf{p}^*(\mathbf{Z}) \equiv \mathbf{p}(\mathbf{Z}, \hat{\boldsymbol{\beta}}_1)$ is preferable to another predictor $\mathbf{p}^{**}(\mathbf{Z}) \equiv \mathbf{p}(\mathbf{Z}, \hat{\boldsymbol{\beta}}_2)$ if $r(\mathbf{p}^*, \mathbf{Z}_0; \boldsymbol{\beta}, \sigma^2) \leqslant r(\mathbf{p}^{**}, \mathbf{Z}_0; \boldsymbol{\beta}, \sigma^2)$, for all $(\boldsymbol{\beta}', \sigma^2) \in R^p x(0, \infty)$.

Assuming the decomposition (3.2) holds, $\mathbf{p}^*$ is preferable to $\mathbf{p}^{**}$ if and only if

$$\sigma^{-2}E\{(\hat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta})' Q(\hat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta})\} - \sigma^{-2}E\{(\hat{\boldsymbol{\beta}}_2 - \boldsymbol{\beta})' Q(\hat{\boldsymbol{\beta}}_2 - \boldsymbol{\beta})\} \leqslant 0$$
$$\text{for all} \quad (\boldsymbol{\beta}', \sigma^2) \in R^p x(0, \infty); \quad (3.4)$$

i.e., if and only if $\hat{\boldsymbol{\beta}}_1$ has a uniformly smaller (or equal) weighted mean squared error than that of $\hat{\boldsymbol{\beta}}_2$. Thus, the problem of *prediction* and the comparison of *predictors* has been reduced to the problem of *estimation* and the comparison of *estimators*.

4. IMPROVED PREDICTION

In this section we construct predictors, based on the general class (2.6), that have risks less than or equal to that of the best linear unbiased predictor, over regions of $(\boldsymbol{\beta}', \sigma^2) \in R^p x(0, \infty)$. The minimax predictor to be defined in Section 4.3 is shown to achieve *uniform* improvement of risk.

4.1. *Bayesian Prediction*

In model (1.1), suppose

$$\mathbf{Z} \mid \boldsymbol{\beta} \sim N(X\boldsymbol{\beta}, \sigma^2\Sigma_{ZZ}) \qquad \text{and} \qquad \boldsymbol{\beta} \sim N(\boldsymbol{\theta}, \Sigma_{\beta\beta}). \quad (4.1)$$

The distribution of $\boldsymbol{\beta}$ reflects the prior information about the unknown parameter $\boldsymbol{\beta}$. Here σ^2 is a nuisance parameter assumed fixed but unknown; a prior for σ^2 could also have been chosen, but this will not change the estimation of $\boldsymbol{\beta}$. Straightforward algebra (see, e.g., [5]) gives

$$\hat{\boldsymbol{\beta}}_{\mathbf{B}} \equiv (X'\Sigma_{ZZ}^{-1}X + \Sigma_{\beta\beta}^{-1})^{-1}(\Sigma_{\beta\beta}^{-1}\boldsymbol{\theta} + X'\Sigma_{ZZ}^{-1}\mathbf{Z}) \tag{4.2}$$

as the Bayes estimator of $\boldsymbol{\beta}$. Substituting (4.2) into (2.6) gives the Bayes predictor of $\mathbf{Z}_0$:

$$\begin{aligned}\mathbf{p}_5(\mathbf{Z}) &\equiv \mathbf{p}(\mathbf{Z}, \hat{\boldsymbol{\beta}}_{\mathbf{B}})\\ &= \Sigma_{0Z}\Sigma_{ZZ}^{-1}\mathbf{Z} + (X_0 - \Sigma_{0Z}\Sigma_{ZZ}^{-1}X)(X'\Sigma_{ZZ}^{-1}X + \Sigma_{\beta\beta}^{-1})^{-1}\\ &\quad\times(\Sigma_{\beta\beta}^{-1}\boldsymbol{\theta} + X'\Sigma_{ZZ}^{-1}\mathbf{Z}).\end{aligned} \tag{4.3}$$

This is precisely the Bayes predictor derived by Kitanidis [18]. Note that when $\Sigma_{\beta\beta}^{-1} = \phi$ (a $p \times p$ matrix of zeros), corresponding to no information on $\boldsymbol{\beta}$, Eq. (4.3) reduces to $\mathbf{p}_2(\mathbf{Z})$ of (2.8), the best linear unbiased (kriging) predictor. Also, taking $\boldsymbol{\theta} = \mathbf{0}$ and $\Sigma_{\beta\beta} = (1/\kappa) I_p$, for some constant κ in (4.2), gives the ridge-regression estimator

$$\hat{\boldsymbol{\beta}}_{\mathbf{R}} \equiv (X'\Sigma_{ZZ}^{-1}X + \kappa I_p)^{-1}(X'\Sigma_{ZZ}^{-1}\mathbf{Z}) \tag{4.4}$$

[14] and hence a ridge predictor

$$\begin{aligned}\mathbf{p}_6(\mathbf{Z}) &\equiv \mathbf{p}(\mathbf{Z}, \hat{\boldsymbol{\beta}}_{\mathbf{R}})\\ &= \Sigma_{0Z}\Sigma_{ZZ}^{-1}\mathbf{Z} + (X_0 - \Sigma_{0Z}\Sigma_{ZZ}^{-1}X)(X'\Sigma_{ZZ}^{-1}X + \kappa I_p)^{-1}(X'\Sigma_{ZZ}^{-1}\mathbf{Z}).\end{aligned} \tag{4.5}$$

Empirical Bayes predictors may also be considered by estimating $\boldsymbol{\theta}$ in (4.3) with an estimator $\hat{\boldsymbol{\theta}}$.

The Bayes predictor (4.3) is biased:

$$\begin{aligned}E(\mathbf{p}_5(\mathbf{Z}) - \mathbf{Z}_0 \mid \boldsymbol{\beta}) &= (\Sigma_{0Z}\Sigma_{ZZ}^{-1}X - X_0)\boldsymbol{\beta} + (X_0 - \Sigma_{0Z}\Sigma_{ZZ}^{-1}X)\\ &\quad\times(X'\Sigma_{ZZ}^{-1}(X'\Sigma_{ZZ}^{-1}X + \Sigma_{\beta\beta}^{-1})^{-1}\boldsymbol{\theta} + X'\Sigma^{-1}X\boldsymbol{\beta}).\end{aligned}$$

Also, since $\hat{\boldsymbol{\beta}}_{\mathbf{B}}$ is linear in the data, Lemma 3.1 applies and from (3.2) its risk (conditional on $\boldsymbol{\beta}$) is

$$\begin{aligned}r(\mathbf{p}_5, \mathbf{Z}_0; \boldsymbol{\beta}, \sigma^2) &= \mathrm{tr}(\Sigma_{00} - \Sigma_{0Z}\Sigma_{ZZ}^{-1}\Sigma_{Z0}) + \mathrm{tr}(T'QT(X'\Sigma_{ZZ}^{-1}X)^{-1})\\ &\quad+ \sigma^{-2}(\boldsymbol{\beta} - \boldsymbol{\theta})' W'QW(\boldsymbol{\beta} - \boldsymbol{\theta}),\end{aligned} \tag{4.6}$$

where $T=(X'\Sigma_{ZZ}^{-1}X+\Sigma_{\beta\beta}^{-1})^{-1}(X'\Sigma_{ZZ}^{-1}X)$, $W=I-T$, and Q is given by (3.3). From this we can see that $r(\mathbf{p}_5, \mathbf{Z}_0; \boldsymbol{\beta}, \sigma^2) \leqslant r(\mathbf{p}_2, \mathbf{Z}_0; \boldsymbol{\beta}, \sigma^2)$ if and only if

$$\sigma^{-2}(\boldsymbol{\beta}-\boldsymbol{\theta})' W'QW(\boldsymbol{\beta}-\boldsymbol{\theta}) \leqslant \operatorname{tr}(X'\Sigma_{ZZ}^{-1}X)^{-1} - \operatorname{tr}(T'QT(X'\Sigma_{ZZ}^{-1}X)^{-1}). \tag{4.7}$$

It is possible to satisfy (4.7) if $\boldsymbol{\beta}$ is close to $\boldsymbol{\theta}$. Of course, the problem is that $\boldsymbol{\beta}$ is unknown. Hoerl and Kennard [14] give some discussion on this problem in the context of ridge regression. For further discussion, see Smith [21] and Giles and Rayner [11].

4.2. *Minimax Prediction*

Judge and Bock [16] consider the family of estimators of $\boldsymbol{\beta}$:

$$\hat{\boldsymbol{\beta}}_{\mathrm{JB}} = \{I_p - h(\hat{\boldsymbol{\beta}}'_{\mathrm{GLS}} B \hat{\boldsymbol{\beta}}_{\mathrm{GLS}}/(n-p)\,\hat{\sigma}^2)\, C\}\, \hat{\boldsymbol{\beta}}_{\mathrm{GLS}}, \tag{4.8}$$

where

B and C are $p \times p$ real matrices;
$h(\cdot)$ is a real-valued function;
$\hat{\sigma}^2 = \{1/(n-p)\}(\mathbf{Z}-X\hat{\boldsymbol{\beta}}_{\mathrm{GLS}})'\,\Sigma_{ZZ}^{-1}(\mathbf{Z}-X\hat{\boldsymbol{\beta}}_{\mathrm{GLS}})$;
$\hat{\boldsymbol{\beta}}_{\mathrm{GLS}}$ is the generalized least squares estimator of $\boldsymbol{\beta}$ given in Eq. (2.7), and I_p is the $p \times p$ identity matrix. (4.9)

Using $\hat{\boldsymbol{\beta}}_{\mathrm{JB}}$ as an estimator of $\boldsymbol{\beta}$ in (2.6), we obtain the corresponding family of predictors

$$\mathbf{p}_7(\mathbf{Z}) \equiv \mathbf{p}(\mathbf{Z}, \hat{\boldsymbol{\beta}}_{\mathrm{JB}}) = \Sigma_{0Z}\Sigma_{ZZ}^{-1}\mathbf{Z} + (X_0 - \Sigma_{0Z}\Sigma_{ZZ}^{-1}X)\,\hat{\boldsymbol{\beta}}_{\mathrm{JB}}. \tag{4.10}$$

The conditions under which this family of predictors has risk (given by (3.2)) less than or equal to that of the best linear unbiased predictor are given in the following theorem.

THEOREM 4.1. *Assume* $(\mathbf{Z}', \mathbf{Z}_0')'$ *are jointly Gaussian with mean and covariance given by* (1.1) *and* (1.2), *respectively. Furthermore, assume that* Q *given by* (3.3) *has rank* p, *and matrices* C *and* B *of* (4.8) *are chosen so that* $Q^{1/2}CQ^{-1/2}$ *and* $Q^{1/2}BQ^{-1/2}$ *are positive-definite matrices that commute with each other and that also commute with* $Q^{1/2}(X'\Sigma_{ZZ}^{-1}X)^{-1}Q^{1/2}$. *Let* $\lambda_1(D)$ *and* $\operatorname{tr}(D)$ *denote the maximum eigenvalue and the trace of any square matrix* D. *If*

$$\text{(i)}\quad 0 \leqslant c \leqslant \frac{2\{\operatorname{tr}(C(X'\Sigma_{ZZ}^{-1}X)^{-1}Q - 2\lambda_1(C(X'\Sigma_{ZZ}^{-1}X)^{-1}Q))\}}{(n-p+2)\lambda_1(C'QCB^{-1})}; \qquad (4.11)$$

(ii) $0 \leqslant h(u) \leqslant c/u$, *for all* $u > 0$, *and* h *is differentiable for all* $u > 0$;

(iii) $\psi(u) \equiv u^q\{(c/u) - h(u)\}^{-(1+f)} h(u)$

is nondecreasing in u *if* $h(u) \leqslant c/u$, *where*

$$q = \{c(n-p-2)/4\}\{\lambda_1(C'QCB^{-1})/\lambda_1(C(X'\Sigma_{ZZ}^{-1}X)^{-1}Q)\}$$

and

$$f = (4/(n-p-2))\,q;$$

then

$$r(\mathbf{p}_7, \mathbf{Z}_0; \boldsymbol{\beta}, \sigma^2) \leqslant r(\mathbf{p}_2, \mathbf{Z}_0; \boldsymbol{\beta}, \sigma^2) \qquad \textit{for all} \quad (\boldsymbol{\beta}', \sigma^2) \in R^p x(0, \infty).$$

Here, $r(\cdot, \mathbf{Z}_0; \boldsymbol{\beta}, \sigma^2)$ *is given by* (3.2), $\mathbf{p}_2(\mathbf{Z})$ *is the best linear unbiased predictor of* $\mathbf{Z}_0$ *given by* (2.8), *and* $\mathbf{p}_7$ *is given by* (4.10).

Proof. Since $(\mathbf{Z}', \mathbf{Z}_0')$ are jointly Gaussian, Lemma 3.2 applies, so that from (3.2),

$$r(\mathbf{p}_7, \mathbf{Z}_0; \boldsymbol{\beta}, \sigma^2) = \operatorname{tr}(\Sigma_{00} - \Sigma_{0Z}\Sigma_{ZZ}^{-1}\Sigma_{Z0}) + \sigma^2 E(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})' Q(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}).$$

Substituting in for $\mathbf{p}_7(\mathbf{Z})$, $\mathbf{p}_2(\mathbf{Z})$, $\hat{\boldsymbol{\beta}}_{JB}$, and $\hat{\boldsymbol{\beta}}_{GLS}$ using (4.10), (2.8), (4.8), and (2.7), respectively, gives

$$\begin{aligned} &r(\mathbf{p}_7, \mathbf{Z}_0; \boldsymbol{\beta}, \sigma^2) - r(\mathbf{p}_2, \mathbf{Z}_0; \boldsymbol{\beta}, \sigma^2) \\ &\quad = \sigma^{-2}\{E((\hat{\boldsymbol{\beta}}_{JB} - \boldsymbol{\beta})' Q(\hat{\boldsymbol{\beta}}_{JB} - \boldsymbol{\beta})) - E((\hat{\boldsymbol{\beta}}_{GLS} - \boldsymbol{\beta})' Q(\hat{\boldsymbol{\beta}}_{GLS} - \boldsymbol{\beta}))\}. \end{aligned}$$

Conditions (i), (ii), and (iii) ensure that $\hat{\boldsymbol{\beta}}_{JB}$ is minimax [16, p. 234] under loss function $\sigma^{-2}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})' Q(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})$. Hence

$$r(\mathbf{p}_7, \mathbf{Z}_0; \boldsymbol{\beta}, \sigma^2) \leqslant r(\mathbf{p}_2, \mathbf{Z}_0; \boldsymbol{\beta}, \sigma^2) \qquad \text{for all} \quad (\boldsymbol{\beta}', \sigma^2) \in R^p x(0, \infty).$$

Theorem 4.1 shows that a large class of (nonlinear) predictors can be found that have uniformly smaller or equal risk than the best linear unbiased (kriging) predictor.

4.3. *Improved Nonlinear Prediction*

COROLLARY 4.1. *Under the assumptions and notation of Theorem* 4.1, *if*

$$\mathbf{p}_8(\mathbf{Z}) \equiv \mathbf{p}(\mathbf{Z}, \hat{\boldsymbol{\beta}}_{GJS}) = \Sigma_{0Z}\Sigma_{ZZ}^{-1}\mathbf{Z} + (X_0 - \Sigma_{0Z}\Sigma_{ZZ}^{-1}X)\,\hat{\boldsymbol{\beta}}_{GJS}, \qquad (4.12)$$

where

$$\hat{\boldsymbol{\beta}}_{\text{GJS}} = \left(1 - \frac{a(n-p)\,\hat{\sigma}^2}{\hat{\boldsymbol{\beta}}'_{\text{GLS}} X' \Sigma_{ZZ}^{-1} X \hat{\boldsymbol{\beta}}_{\text{GLS}}}\right) \hat{\boldsymbol{\beta}}_{\text{GLS}}, \tag{4.13}$$

$$0 \leqslant a \leqslant \frac{2[\operatorname{tr}\{(X'\Sigma_{ZZ}^{-1}X)^{-1}Q\}\,\lambda_1^{-1} - 2]}{(n-p+2)}, \tag{4.14}$$

and λ_1 *is the largest eigenvalue of* $(X'\Sigma_{ZZ}^{-1}X)^{-1}Q$, *then*

$$r(\mathbf{p}_8, \mathbf{Z}_0; \boldsymbol{\beta}, \sigma^2) \leqslant r(\mathbf{p}_2, \mathbf{Z}_0; \boldsymbol{\beta}, \sigma^2) \qquad \textit{for all} \quad (\boldsymbol{\beta}', \sigma^2) \in R^p x(0, \infty).$$

Proof. Take $h(x) = a/x$, $a \in R$, $B = (X'\Sigma_{ZZ}^{-1}X)$, and $C = I_p$ and apply Theorem 4.1. Note that in the case of James–Stein estimation, the Gaussianity assumption may be relaxed to include elliptically symmetric distributions [22]. As a consequence (see Lemma 3.2), the uniform reduction in risk over that of the BLUP also holds for elliptically symmetric distributions as well.

The condition (4.14) of Corollary 4.1 warrants some discussion. The assumption that Q is full rank p implies that $k \geqslant p$, so that at least p random variables must be predicted. Also, p (the dimensionality of the estimation problem) must be larger than two [15]. If $(X'\Sigma_{ZZ}^{-1}X)$ is ill-conditioned, (4.14) may fail to hold. The matrix Q depends on the relationship of prediction locations to data locations and requires that the two be reasonably close together. This suggests that (4.14) may be more likely to hold for interpolation problems than for extrapolation situations.

For the interval defined by (4.14) to exist, $\operatorname{tr}\{(X'\Sigma_{ZZ}^{-1}X)^{-1}Q\}\,\lambda_1^{-1}$ must be at least two. But $\operatorname{tr}\{(X'\Sigma_{ZZ}^{-1}X)^{-1}Q\}\,\lambda_1^{-1} = \Sigma\lambda_i/\max\{\lambda_i\}$, where the λ_i's are the eigenvalues of $(X'\Sigma_{ZZ}^{-1}X)^{-1}Q$. In a spatial setting, a natural design criterion for the choice of prediction locations is the maximization of $\Sigma\lambda_i/\max\{\lambda_i\}$. There are probably many choices of X and X_0, particularly in the spatial and temporal settings, for which there is no interval defined by (4.14). In this case, no predictor of the general form of $\mathbf{p}_8(\mathbf{Z})$ will be minimax.

Expressions for the bias and the risk may be derived for any member of the class $\mathbf{p}_7(\mathbf{Z})$ (and hence for $\mathbf{p}_8(\mathbf{Z})$) once $h(\cdot)$ has been specified. Straightforward extensions of theorems in Appendix B of Judge and Bock [16] give

$$E(\mathbf{p}_8(\mathbf{Z}) - \mathbf{Z}_0) = -a(X_0 - \Sigma_{0Z}\Sigma_{ZZ}^{-1}X)\, E(\chi^2_{(n-p)}/\chi^2_{(p+2,\lambda)})\, \boldsymbol{\beta}, \tag{4.15}$$

and

$$\begin{aligned} r(\mathbf{p}_8, \mathbf{Z}_0; \boldsymbol{\beta}, \sigma^2) = {} & \operatorname{tr}(\Sigma_{00} - \Sigma_{0Z}\Sigma_{ZZ}^{-1}\Sigma_{Z0}) \\ & + \operatorname{tr}(Q(X'\Sigma_{ZZ}^{-1}X)^{-1})\, E\{1 - a\chi^2_{(n-p)}/\chi^2_{(p+2,\lambda)}\}^2 \\ & + \boldsymbol{\beta}'Q\boldsymbol{\beta}\{E(1 - \chi^2_{(n-p)}/\chi^2_{(p+4,\lambda)})^2 \\ & + 2E(\chi^2_{(n-p)}/\chi^2_{(p+2,\lambda)}) - 1\}, \end{aligned} \tag{4.16}$$

where $\chi^2_{(v,\lambda)}$ is a noncentral chi-squared random variable on v degrees of freedom, with noncentrality parameter $\lambda = \boldsymbol{\beta}' X' \Sigma_{ZZ}^{-1} X \boldsymbol{\beta} / 2\sigma^2$, independent of the central chi-squared random variable $\chi^2_{(n-p)}$.

The expressions (4.15) and (4.16) may be evaluated using the computational formulas for inverse moments of a noncentral chi-squared distribution given in Bock *et al.* [6] and Xie [26]. Patnaik [19] shows that a chi-squared distribution can be closely approximated by a scalar multiple of a central chi-squared random variable, which could be used to simplify (4.15) and (4.16). Of course, all expressions depend on the unknowns $\boldsymbol{\beta}$ and σ^2, but estimators based on $\hat{\boldsymbol{\beta}}_{\mathrm{GLS}}$ and $\hat{\sigma}^2$ can be used.

It is known that the positive part of $\hat{\boldsymbol{\beta}}_{\mathrm{GJS}}$ [2],

$$\hat{\boldsymbol{\beta}}^+_{\mathrm{GJS}} = \left(1 - \frac{a(n-p)\,\hat{\sigma}^2}{\hat{\boldsymbol{\beta}}'_{\mathrm{GLS}} X' \Sigma_{ZZ}^{-1} X \hat{\boldsymbol{\beta}}_{\mathrm{GLS}}}\right)^+ \hat{\boldsymbol{\beta}}_{\mathrm{GLS}}, \tag{4.17}$$

where $g^+ \equiv \max\{0, g\}$, dominates $\hat{\boldsymbol{\beta}}_{\mathrm{GJS}}$. That is,

$$E((\hat{\boldsymbol{\beta}}^+_{\mathrm{GJS}} - \boldsymbol{\beta})' W (\hat{\boldsymbol{\beta}}^+_{\mathrm{GJS}} - \boldsymbol{\beta})) \leqslant E((\hat{\boldsymbol{\beta}}_{\mathrm{GJS}} - \boldsymbol{\beta})' W (\hat{\boldsymbol{\beta}}_{\mathrm{GJS}} - \boldsymbol{\beta})), \tag{4.18}$$

for all $(\boldsymbol{\beta}', \sigma^2) \in R^p x(0, \infty)$, and any nonnegative-definite weight matrix W; for a proof of this see Judge and Bock [16, p. 239]. Consequently, with Q playing the role of W in (4.18), the predictor

$$\mathbf{p}_9(\mathbf{Z}) \equiv \mathbf{p}(\mathbf{Z}, \hat{\boldsymbol{\beta}}^+_{\mathrm{GJS}}) = \Sigma_{0Z} \Sigma_{ZZ}^{-1} \mathbf{Z} + (X_0 - \Sigma_{0Z} \Sigma_{ZZ}^{-1} X)\, \hat{\boldsymbol{\beta}}^+_{\mathrm{GJS}} \tag{4.19}$$

must necessarily have smaller risk (as given in (1.5)) than that of $\mathbf{p}_8(\mathbf{Z})$ in (4.15).

Expressions for the risk of the estimator $\hat{\boldsymbol{\beta}}^+_{\mathrm{GJS}}$, and hence for the risk of the predictor, $\mathbf{p}_9(\mathbf{Z})$, are intractable. However, (4.16) could be used instead, since it would give a conservative value of the risk of $\mathbf{p}_9(\mathbf{Z})$.

The predictor $\mathbf{p}_7$ has a nice interpretation as a shrinkage predictor. Since $\hat{\boldsymbol{\beta}}_{\mathrm{GJS}}$ shrinks the generalized least squares estimator $\hat{\boldsymbol{\beta}}_{\mathrm{GLS}}$ towards zero, the corresponding predictor in (4.10) shrinks the best linear unbiased predictor $\mathbf{p}_2(\mathbf{Z})$ to the best linear predictor under the assumption that $\boldsymbol{\beta}$ is known and equal to $\mathbf{0}$; i.e.,

$$\mathbf{p}_7(\mathbf{Z}) = \{I_p - h(\hat{\boldsymbol{\beta}}'_{\mathrm{GLS}} B \hat{\boldsymbol{\beta}}_{\mathrm{GLS}} / (n-p)\,\hat{\sigma}^2)\, C\}(\mathbf{p}_2(\mathbf{Z}) - \Sigma_{0Z} \Sigma_{ZZ}^{-1} \mathbf{Z}) + \Sigma_{0Z} \Sigma_{ZZ}^{-1} \mathbf{Z}.$$

Throughout this development of minimax prediction, we have chosen to focus on the family of minimax estimators developed by Judge and Bock [16], simply because the condition for minimaxity are concise and easy to use, and because the familiar James–Stein type estimators are members of this class. However, our results are not limited to the consideration of this one class. Other minimax predictors can be constructed using the minimax estimators developed by, e.g., Berger [3] and Strawderman [23], or any in the family of admissible estimators developed by Alam [1]. In practical situations, the choice of the estimator (and hence the predictor) is a difficult one; see Berger [4] for some interesting ideas on minimax estimator selection.

If the condition on the constant a in (4.14) is not met, we recommend using the universal kriging predictor given in (2.8). However, if this condition is satisfied, then the predictor $\mathbf{p}_9(\mathbf{Z})$ has risk which is uniformly smaller than that of the best linear unbiased predictor and is the predictor of choice among those presented in this paper.

5. Example

In this section, we compare the risk of three of the predictors presented in this manuscript, based on the small spatial data configuration illustrated in Fig. 1. In order to make this comparison, the linear model given by Eq. (1.1) must be completely specified. This will be done using geostatistical ideas.

Denote a typical spatial location in Fig. 1 by $\mathbf{s} = (x, y)'$, and the datum associated with this location as $Z(\mathbf{s})$. Then, taking the rows of the matrices, X and X_0 to be linear functions of the spatial indices, the linear model of Eq. (1.1) may be written as

$$\mathbf{Z} = \begin{bmatrix} Z(0.0, 0.0) \\ Z(1.0, 0.0) \\ Z(2.0, 0.0) \\ Z(0.0, 1.0) \\ Z(1.0, 1.0) \\ Z(2.0, 1.0) \\ Z(0.0, 2.0) \\ Z(1.0, 2.0) \\ Z(2.0, 2.0) \end{bmatrix} = \begin{bmatrix} 1.0 & 0.0 & 0.0 \\ 1.0 & 1.0 & 0.0 \\ 1.0 & 2.0 & 0.0 \\ 1.0 & 0.0 & 1.0 \\ 1.0 & 1.0 & 1.0 \\ 1.0 & 2.0 & 1.0 \\ 1.0 & 0.0 & 2.0 \\ 1.0 & 1.0 & 2.0 \\ 1.0 & 2.0 & 2.0 \end{bmatrix} \boldsymbol{\beta} + \boldsymbol{\varepsilon} \tag{5.1}$$

$$\mathbf{Z}_0 = \begin{bmatrix} Z(0.5, 0.5) \\ Z(1.0, 0.5) \\ Z(1.5, 0.5) \\ Z(0.5, 1.5) \\ Z(1.5, 1.5) \end{bmatrix} = \begin{bmatrix} 1.0 & 0.5 & 0.5 \\ 1.0 & 1.0 & 0.5 \\ 1.0 & 1.5 & 0.5 \\ 1.0 & 0.5 & 1.5 \\ 1.0 & 1.5 & 1.5 \end{bmatrix} \boldsymbol{\beta} + \boldsymbol{\varepsilon}_0.$$

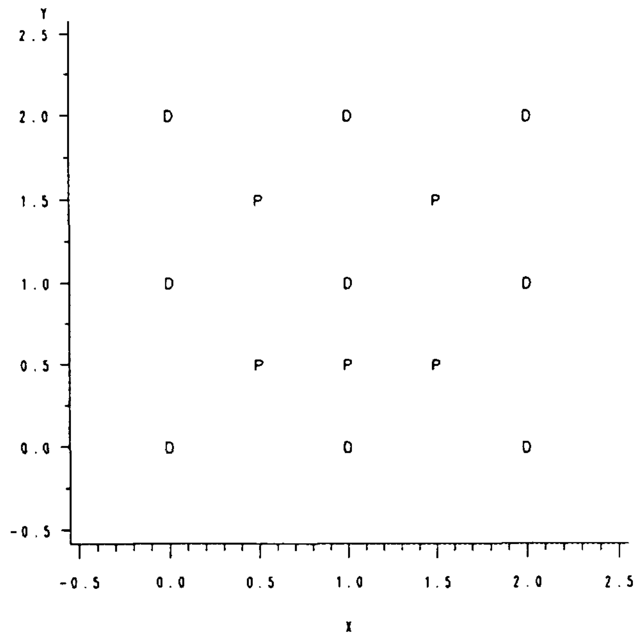

FIG. 1. Spatial data configuration for the example presented in Section 5; D denotes data locations and P denotes the locations of values to be predicted from the data.

In a geostatistical framework, the covariance structure of the data is specified as a function of the distance between spatial locations. In this example, it is derived from

$$C(\|\mathbf{h}\|) = \begin{cases} 1.0 - 1.5(\|\mathbf{h}\|/1.5) + 0.5(\|\mathbf{h}\|/1.5)^3, & \|\mathbf{h}\| \leqslant 1.5; \\ 0, & \|\mathbf{h}\| > 1.5, \end{cases} \tag{5.2}$$

where $\|\mathbf{h}\|$ is the Euclidean distance between spatial locations separated by the vector $\mathbf{h}$, and $C(0) \equiv \sigma^2 = 1$. Then $\text{cov}(Z(\mathbf{s}), Z(\mathbf{u})) = C(\|\mathbf{s} - \mathbf{u}\|)$; for example, $\text{cov}(Z(1.0, 1.0), Z(2.0, 1.0)) = C(1.0) = 0.15$, and so the $(5, 6)$th element of Σ_{ZZ} is 0.15. The only unknown in the linear model of Eq. (1.1) is the 3×1 vector $\boldsymbol{\beta}$. The risk of the predictors will be computed and compared for various values of $\boldsymbol{\beta}$.

Three predictors were investigated in this example: the ridge predictor, $\mathbf{p}_6(\mathbf{Z})$, of Eq. (4.5), the James–Stein predictor, $\mathbf{p}_8(\mathbf{Z})$, of Eq. (4.12), and the best linear unbiased predictor (BLUP), $\mathbf{p}_2(\mathbf{Z})$, of Eq. (2.8). For each predictor, expressions for the risk were evaluated for various $\boldsymbol{\beta}$ vectors. Clearly, since the risk of the (BLUP) does not depend on $\boldsymbol{\beta}$, it is constant.

The risk of the ridge predictor is given by Eq. (4.6) using $\theta = 0$ and $\Sigma_{\beta\beta} = (1/\kappa) I_3$. Since the length of the corresponding ridge estimator of

Eq. (4.4) tends to zero as κ approaches infinity, and for fixed β, the risk of the ridge estimator decreases with increasing κ, large values of κ were chosen for this example. Results with $\kappa = 50$ are presented here.

The risk of the James–Stein predictor, given by Eq. (4.16), is more complicated. In computing this quantity, a value of the constant, a, must be specified. For this example, a was taken to be the upper bound in condition (4.14), so that $a = 2[\mathrm{tr}\{(X'\Sigma_{ZZ}^{-1}X)^{-1}Q\}\lambda_1^{-1} - 2]/(n-p+2)$. To evaluate Eq. (4.16) for different $\boldsymbol{\beta}$ vectors, the expectations of ratios of noncentral chi-squared random variables are needed. These were obtained using the method given in Xie [26]. In general, this method is very easy to use; however, caution should be taken with $\boldsymbol{\beta}$ close to $\mathbf{0}$, since the distribution of the noncentral chi-squared random variable is nearly that of a central chi-squared random variable.

Finally, the risk of the BLUP, given just below (2.8), is constant for this example and is equal to $\mathrm{tr}\{(X'\Sigma_{ZZ}^{-1}X)^{-1}Q\}$, where Q is given by Eq. (3.3).

The expressions for the risk associated with each of the three predictors were evaluated for 40 different $\boldsymbol{\beta}$ vectors with Euclidean norm ranging from 0.0 to 4.0. Fig. 2 shows the risk for the James–Stein predictor relative to that for the BLUP as a function of the norm of $\boldsymbol{\beta}$. Specifically, the vertical axis represents $\{r(\mathbf{p}_8, \mathbf{Z}_0; \boldsymbol{\beta}, \sigma^2)/r(\mathbf{p}_2, \mathbf{Z}_0; \boldsymbol{\beta}, \sigma^2)\} \times 100$ is plotted against

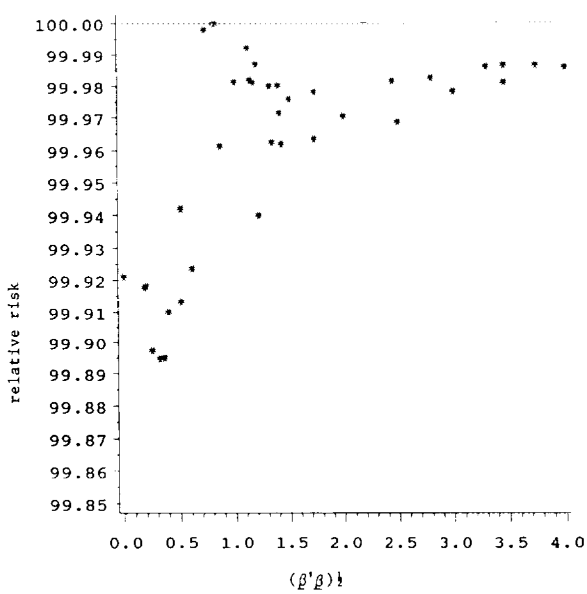

FIG. 2. Relative risk of the James–Stein predictor as a function of $(\boldsymbol{\beta}'\boldsymbol{\beta})^{1/2}$. The graph is based on the spatial configuration of Fig. 1 and the example described in Section 5 of the text. The vertical axis is computed as $\{r(\mathbf{p}_8, \mathbf{Z}_0; \boldsymbol{\beta})/r(\mathbf{p}_2, \mathbf{Z}_0; \boldsymbol{\beta})\} \times 100$. The horizontal axis is $(\boldsymbol{\beta}'\boldsymbol{\beta})^{1/2}$, the Euclidean norm of $\boldsymbol{\beta}$.

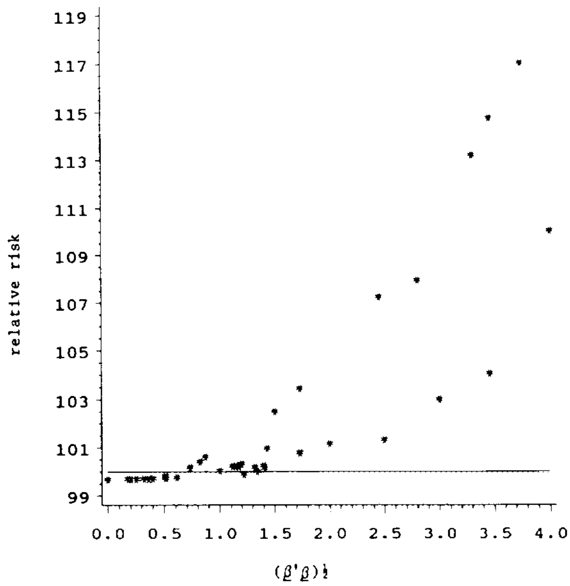

FIG. 3. Relative risk of the ridge predictor as a function of the $(\boldsymbol{\beta}'\boldsymbol{\beta})^{1/2}$. The graph is based on the spatial configuration of Fig. 1 and the example described in Section 5 of the text. The vertical axis is computed as $\{r(\mathbf{p}_6, \mathbf{Z}_0; \boldsymbol{\beta})/r(\mathbf{p}_2, \mathbf{Z}_0; \boldsymbol{\beta})\} \times 100$. The horizontal axis is $(\boldsymbol{\beta}'\boldsymbol{\beta})^{1/2}$, the Euclidean norm of $\boldsymbol{\beta}$.

$(\boldsymbol{\beta}'\boldsymbol{\beta})^{1/2}$. The uniform reduction in risk using the James–Stein predictor is readily apparent, and the greatest reduction in risk occurs for $\boldsymbol{\beta}$ near $\mathbf{0}$, as is expected from similar results on James–Stein estimation. Figure 3 shows the relative risk for the ridge predictor. The strong increase in risk with increasing $\boldsymbol{\beta}'\boldsymbol{\beta}$, is consistent with results for ridge estimation, again indicating that the predictor tends to reflect the properties of its associated estimator. Note that for many $\boldsymbol{\beta}$, the risk associated with the ridge predictor is much greater than that of the BLUP.

The results of this section clearly illustrate the uniform reduction in risk that can be obtained using the James–Stein predictor of Eq. (4.12). Although the reduction attained in this example was not very large, larger risk reductions may be attainable for different spatial data configurations.

ACKNOWLEDGMENT

This research was partially supported by the National Science Foundation under Grants DMS-8703083 and DMS-8902812.

REFERENCES

[1] ALAM, K. (1973). A family of admissible estimators for the mean of a multivariate normal distribution. *Ann. Statist.* **1** 517–525.

[2] BARANCHIK, A. J. (1964). *Multiple Regression and Estimation of the Mean of a Multivariate Normal Distribution. Technical Report No. 51.* Stanford University, Stanford, CA.

[3] BERGER, J. O. (1976). Admissible minimax estimation of a multivariate normal mean with arbitrary quadratic loss. *Ann. Statist.* **4** 223–226.

[4] BERGER, J. O. (1982). Selecting a minimax estimator of a multivariate normal mean. *Ann. Statist.* **10** 81–92.

[5] BERGER, J. O. (1985). *Statistical Decision Theory and Bayesian Analysis*, 2nd ed. Springer-Verlag, New York.

[6] BOCK, M. E., JUDGE, G. G., AND YANCEY, T. A. (1984). A simple form for the inverse moments of non-central χ^2 and F random variables and certain confluent hypergeometric functions. *J. Econometrics* **25** 217–234.

[7] COPAS, J. B. (1983). Regression, prediction, and shrinkage. *J. Roy. Stat. Soc. Ser. B* **45** 311–354.

[8] COPAS, J. B., AND JONES, M. C. (1987). Regression shrinkage methods and autoregressive time series prediction. *Austral. J. Statist.* **29** 264–277.

[9] EFRON, B., AND MORRIS, C. (1973). Stein's estimation rule and its competitors—An empirical Bayes approach. *J. Amer. Statist. Assoc.* **68** 117–130.

[10] FULLER, W. A. (1976). *Introduction to Statistical Time Series.* Wiley, New York.

[11] GILES, D. E. A., AND RAYNER, A. C. (1979). The mean squared errors of the maximum likelihood and natural-conjugate Bayes regression estimators. *J. Econometrics* **11** 319–334.

[12] GOLDBERGER, A. S. (1962). Best linear unbiased prediction in the generalized linear regression model. *J. Amer. Statist. Assoc.* **57** 369–375.

[13] HARVILLE, D. A. (1985). Decomposition of prediction error. *J. Amer. Statist. Assoc.* **80** 132–138.

[14] HOERL, A. E., AND KENNARD, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics* **12** 55–67.

[15] JAMES, W., AND STEIN, C. (1961). Estimation with quadratic loss. In *Proceedings of the Fourth Berkeley Symposium on Mathematics, Statistics and Probability* (Jerzy Neyman, Ed.). Vol. I, pp. 361–379. Univ. of California Press, Berkeley.

[16] JUDGE, G. E., AND BOCK, M. E. (1978). *The Statistical Implications of Pre-Test and Stein-Rule Estimators in Econometrics.* North-Holland, Amsterdam.

[17] KELKER, D. (1970). Distribution theory of spherical distributions and a location-scale parameter generalization. *Sankhyā A* **32** 419–430.

[18] KITANIDIS, P. K. (1986). Parameter uncertainty in estimation of spatial functions: Bayesian analysis. *Water Resources Res.* **22** 499–507.

[19] PATNAIK, P. B. (1949). The non-central X^2- and F-distributions and their applications. *Biometrika* **36** 202–232.

[20] RAO, C. R. (1973). *Linear Statistical Inference.* Wiley, New York.

[21] SMITH, A. F. M. (1973). A general Bayesian linear model. *J. Roy. Statist. Soc. Ser. B* **35** 67–75.

[22] SRIVASTAVA, V. K., AND BILODEAU, M. (1989). Stein estimation under elliptical distributions. *J. Multivariate Anal.* **28** 247–259.

[23] STRAWDERMAN, W. E. (1978). Minimax adaptive generalized ridge regression estimators. *J. Amer. Statist. Assoc.* **73** 623–627.

[24] TOUTENBURG, H. (1982). *Prior Information in Linear Models.* Wiley, New York.

[25] WOLD, H. (1938). *A study in the Analysis of Stationary Time Series.* Almqvist & Wiksells, Uppsula.

[26] XIE, W. Z. (1988). A simple way of computing the inverse moments of a non-central chi-square random variable. *J. Econometrics* **37** 389–393.