The Annals of Statistics
1989, Vol. 17, No. 3, 980–1000

ASYMPTOTIC DISTRIBUTIONS OF MINIMUM NORM QUADRATIC ESTIMATORS OF THE COVARIANCE FUNCTION OF A GAUSSIAN RANDOM FIELD[1]

BY MICHAEL STEIN

The University of Chicago

Consider a continuous Gaussian random field $z(x)$ defined on a compact set $R \subset \mathbb{R}^d$ with covariance function of the form $\operatorname{cov}(z(x), z(x')) = \sum_{i=1}^{k} \theta_i K_i(x, x')$, where the K_i's are specified and $\theta = (\theta_1, \dots, \theta_k)'$ is to be estimated. Let $\{x_l\}_{l=1}^{\infty}$ be a sequence of distinct points in R. Based on $z(x_1), \dots, z(x_N)$, minimum norm quadratic estimation can be used to estimate θ. Suppose $K_1, \dots, K_k$ are compatible covariance functions on R, which means that the Gaussian measures with means zero and covariance functions $K_1, \dots, K_k$ are mutually absolutely continuous. Then, as the number of observations N increases, the minimum norm quadratic estimator of $\sum_{i=1}^{k} \theta_i$ is asymptotically normal with variance of order N^{-1}. The minimum norm quadratic estimator of any other linear combination of the θ_i's converges (in L^2) to some nondegenerate random variable. This limit is the same for any two dense sequence of points in R. Thus, a definition of a minimum norm quadratic estimator of θ when $z(\cdot)$ is observed everywhere in R is obtained.

1. Introduction. Consider a random field $z(\cdot)$ with finite second moments defined on a compact set R in $\mathbb{R}^d$ by

$$z(x) = m(x) + e(x), \tag{1.1}$$

where $m(\cdot)$ is the mean of $z(\cdot)$ and $e(\cdot)$ is a random field with mean zero. Based on observing $z(\cdot)$ at $(x_1, \dots, x_N) \in R$, we wish to predict linear functionals of $z(\cdot)$ defined on R. If we model the mean function by

$$Ez(x) = \beta' f(x), \tag{1.2}$$

where $f(\cdot)$ is a known vector-valued function with q components, β is a vector of unknown regression coefficients and

$$K(x, x') = \operatorname{cov}(z(x), z(x')) \quad \text{for } x, x' \in R$$

is specified, we can obtain the best (minimum prediction error variance) linear unbiased predictor of this linear functional under this model. For example, assuming this model is correct, the best linear unbiased predictor of $z(x)$ based on $Z = (z(x_1), \dots, z(x_n))'$ is [Goldberger (1962)]

$$\left\{\sigma'\Sigma^{-1} + \left(f(x) - F\Sigma^{-1}\sigma\right)'\left(F\Sigma^{-1}F'\right)^{-1}F\Sigma^{-1}\right\}Z,$$

assuming all inverses exist, where $F = (f(x_1), \dots, f(x_N))$, $\sigma = (K(x, x_1), \dots,$

Received January 1987; revised March 1988.
[1]Work supported by an NSF Mathematical Sciences Postdoctoral Research Fellowship. This manuscript was prepared using computer facilities supported in part by NSF Grants DMS-86-01732 and DMS-84-04941 to the Department of Statistics at The University of Chicago.
AMS 1980 *subject classifications*. Primary 62M20; secondary 60G30, 60G60.
Key words and phrases. Equivalence of Gaussian measures, kriging, geostatistics.

$K(x, x_N))'$ and Σ is an $N \times N$ matrix with ijth element $K(x_i, x_j)$. This predictor is known as the kriging predictor in the geostatistical literature and is widely used in mining [Journel and Huijbregts (1978)] and hydrology [Kitanidis (1983)].

Perhaps the main problem in implementing kriging is that $K(\cdot, \cdot)$ is unknown in practice and must be estimated from the observations. We will use a model for the covariance function of the form

$$\text{cov}_\theta(z(x), z(x')) = \sum_{i=1}^{k} \theta_i K_i(x, x') \equiv K_\theta(x, x'), \tag{1.3}$$

where $\theta = (\theta_1, \dots, \theta_k)'$ is a vector of parameters to be estimated. Let V be the correct covariance function, which is not necessarily in the parametric class defined by (1.3). In this paper, we will investigate the situation where $K_1, \dots, K_k$ are what I call compatible on R. In order to define compatibility of covariance functions, recall that corresponding to every measurable function $m(\cdot)$ on R and positive definite function $K(\cdot, \cdot)$ on $R \times R$ there is a unique probability measure of a Gaussian random field on R with mean $m(\cdot)$ and covariance function $K(\cdot, \cdot)$, which we will denote by (m, K). We will say that $K_1(\cdot, \cdot)$ and $K_2(\cdot, \cdot)$ are compatible on R if $(0, K_1)$ and $(0, K_2)$ are equivalent (mutually absolutely continuous with respect to each other), which we will denote by

$$(0, K_1) \equiv (0, K_2).$$

Note that the definition of compatibility of covariance functions does not assume that $z(\cdot)$ is Gaussian. While most of the results in this paper will require that $z(\cdot)$ be Gaussian, the properties of linear predictors discussed in Section 6 do not. Conditions for the equivalence of stationary Gaussian fields are given in one dimension by Ibragimov and Rozanov (1978) and in higher dimensions by Yadrenko (1983). The results presented here can be applied to nonstationary random fields, although conditions for determining the equivalence of nonstationary Gaussian fields would be needed. In particular, it would be worthwhile to extend the results in Yadrenko (1983) to Gaussian intrinsic random functions [Matheron (1973)], which, in geostatistics, are a commonly used class of nonstationary phenomena.

The space of random variables of the form

$$\sum_{i=1}^{n} a_i z(x_i), \tag{1.4}$$

where the a_i's are constant, the x_i's are in R and n is finite, along with their limits in L^2, form a Hilbert space with inner product defined by the second moment. If the mean function $m(\cdot)$ is continuous on R and the covariance function $V(\cdot, \cdot)$ is continuous on $R \times R$, then the Hilbert space is easily seen to be separable. Now, Gaussian measures on separable Hilbert spaces are either equivalent or orthogonal [Kuo (1975), page 125]. We will say that K_1 and K_2 are incompatible if the corresponding Gaussian measures $(0, K_1)$ and $(0, K_2)$ are orthogonal, which we will denote by

$$(0, K_1) \perp (0, K_2).$$

Thus, continuous covariance functions are either compatible or incompatible.

We will now apply some elementary properties of equivalent Gaussian measures to the problem of estimating θ in (1.3) when $K_1, \dots, K_k$ are compatible on R and $z(\cdot)$ is Gaussian. First, for Gaussian measures on an infinitely dimensional Hilbert space, $(0, K) \perp (0, cK)$ if $c \neq 1$ [see Kuo (1975), page 123]. Thus, if $k = 1$ in (1.3) and the model is correct, we would expect to be able to estimate θ_1 consistently as the number of observations in R increases. Generalizing to the case $k > 1$, it is possible to show that if $e'\theta \neq e'\theta^*$, where e is a vector of 1's, then $(0, K_\theta) \perp (0, K_{\theta^*})$. However, if p is a vector not proportional to e, then there exist θ and θ^* such that $p'\theta \neq p'\theta^*$, and $(0, K_\theta) \equiv (0, K_{\theta^*})$ [Stein (1987a)]. It follows that $p'\theta$ cannot be consistently estimated based on observations in R if p is not proportional to e. Thus, $e'\theta$ is the only linear combination of θ we can hope to estimate consistently based on an increasing number of observations in R.

In this paper, we will consider the properties of minimum norm quadratic estimators [Rao (1971, 1972, 1973, 1979)] of θ of Gaussian random fields. These estimators have also been applied to estimation of spatial covariance functions in Kitanidis (1983, 1985), Marshall and Mardia (1985) and Stein (1987a). Minimum norm quadratic estimators (MINQE's) are only defined when the covariance structure is linear in the unknown parameters, which accounts for the model chosen in (1.3). We will examine the properties of a particular type of MINQE known as the MINQE(U, I) as N, the number of observations in R, increases. In contrast, Mardia and Marshall (1984) investigated asymptotic properties of maximum likelihood estimators when, roughly speaking, the distance between neighboring observations remains fixed as N increases so that the observation region grows with N. The MINQE(U, I) is defined in Section 2. It depends on a starting choice for θ which we will denote by α. We assume K_α and K_1 are compatible on R, which will be true if $e'\alpha = 1$ and all of the components of α are positive [follows from Kuo (1975), page 123]. Since the MINQE(U, I) is unchanged when α is multiplied by a scalar, assuming $e'\alpha = 1$ is not really a restriction. We also assume that there exists $c > 0$ and β such that $(\beta' f, cK_\alpha) \equiv (m, V)$, where (m, V) is the true Gaussian measure on R. In Section 3, under these conditions, we show that as $N \to \infty$, the MINQE(U, I) of $e'\theta$ is asymptotically normal with mean c and variance $2c^2/N$. This result is independent of the choice of α. The true measure (m, V) need not have mean $\beta' f(\cdot)$ and covariance $K_\theta(\cdot, \cdot)$ for some β and θ; it merely needs to satisfy $(m, V) \equiv (\beta' f, cK_\alpha)$. We also show that the MINQE(U, I) of $p'\theta$ for any given vector p converges in L^2 to some random variable with respect to (m, V) as $N \to \infty$. If p is not a multiple of e, then the MINQE(U, I) of $p'\theta$ is not consistent; nor, as already noted, can any other possible estimator of $p'\theta$ be consistent. Also, the limit will, in general, depend on both α and (m, V). We see that only if p is a multiple of e does the asymptotic distribution of the MINQE(U, I) of $p'\theta$ obey the usual normal approximations for large sample estimators. In the case where

$(\beta' f, cK_\alpha) \perp (m, V)$ for all c and β, the asymptotic behavior of the MINQE(U, I) is unknown.

In Section 4, we show that this L^2 limit of MINQE(U, I)'s of θ has a direct interpretation as a MINQE(U, I) of θ based on observing the infinite sequence $\{z(x_l)\}_{l=1}^{\infty}$. In Section 5, for $f(\cdot)$ and $K(\cdot, \cdot)$ continuous, we show that the MINQE(U, I) of θ based on $\{z(x_l)\}_{l=1}^{\infty}$ is independent of the actual sequence of points as long as they are dense in R. In Section 6, we show that the results in Section 3 are highly dependent on the assumption that the random field is Gaussian. We will also discuss the connection of these results to the problem of predicting a random field with an estimated covariance function. In particular, Stein (1988) showed that, under appropriate conditions, compatible covariance functions give linear predictions which are asymptotically the same. Thus, considering the results on the MINQE(U, I) obtained here, we see that $e'\theta$, the one linear combination of θ that is consistently estimable, is the one linear combination of θ that can have a nonnegligible asymptotic impact on linear predictions. This result is a somewhat unexpected example of what A. P. Dawid calls Jeffreys's law, which roughly says that predictions based on equivalent prior distributions eventually look nearly identical [Dawid (1984)].

Throughout this paper, we will use the notation that $[v_i]$ represents a column vector with ith element v_i and $[v_{ij}]$ a matrix with ijth element v_{ij}.

2. Minimum norm quadratic estimation. We observe the process $z(\cdot)$ as described in Section 1 at $(x_1, \dots, x_N)$ and wish to estimate the vector θ in (1.3). Let $Z = (z(x_1), \dots, z(x_N))'$ and $F = (f(x_1), \dots, f(x_N))$. We will consider estimating θ using the MINQE(U, I), the minimum norm quadratic estimator that is unbiased (the U) and invariant (the I) with respect to changes in the vector of regression coefficients β. If rank$(F) = r$, we can choose a $(N - r) \times N$ matrix D of full rank whose row space is the orthogonal complement of the row space of F. Define $Y = DZ$, so that Y is a set of contrasts ($EY = 0$ for all β). The MINQE(U, I) can be defined directly in terms of Z, but for the purposes of this paper, it will simplify matters considerably to define the MINQE(U, I) in terms of Y. The fact that this reduction is possible and that the resulting estimator is independent of the particular choice of D is implicit in Rao (1971), page 448. Let $\Psi_l = D[K_l(x_i, x_j)]D'$ and $\Psi(0) = \Sigma\theta_l\Psi_l$. For a given vector $p = (p_1, \dots, p_k)'$, a quadratic unbiased invariant estimator of $p'\theta$ will be of the form $Y'HY$, where

$$\operatorname{tr}(H\Psi_l) = p_l \quad \text{for } l = 1, \dots, k. \tag{2.1}$$

If there exists a matrix H satisfying (2.1), we will say that $p'\theta$ is estimable [Rao (1973), page 305]. The MINQE(U, I) chooses H to minimize the norm $\operatorname{tr}(HT)^2$ for some matrix T subject to (2.1). If $z(\cdot)$ is Gaussian with mean $\beta' f(\cdot)$ and covariance function $K_\theta(\cdot, \cdot)$, then

$$\operatorname{var}(Y'HY) = 2\operatorname{tr}(H\Psi(\theta))^2.$$

Thus, T is taken to be $\Psi(\alpha)$, where α is some a priori choice for θ. Throughout this paper, we will assume

$$\text{(2.2)} \qquad \sum_{l=1}^{k} \alpha_l \left[K_l(x_i, x_j) \right] \text{ is positive definite,}$$

which just says that our a priori choice for the covariance matrix of the observations is nonsingular. Under this assumption, we can choose D so that $\Psi(\alpha) = I$. In this case, if $p'\theta$ is estimable, the MINQE(U, I) of $p'\theta$ is given by

$$p'\left[\operatorname{tr} \Psi_i \Psi_j\right]^{-}\left[Y'\Psi_i Y\right],$$

where $[\operatorname{tr} \Psi_i \Psi_j]^-$ is any generalized inverse of $[\operatorname{tr} \Psi_i \Psi_j]$ [Rao (1973), page 305]. Now $p'\theta$ is estimable for all p if and only if $[\operatorname{tr} \Psi_i \Psi_j]$ is invertible [Rao (1979), page 143], in which case we will say that the MINQE(U, I) of θ exists and is given by

$$\text{(2.3)} \qquad \hat{\theta}(\alpha) = \left[\operatorname{tr} \Psi_i \Psi_j\right]^{-1}\left[Y'\Psi_i Y\right].$$

As part of the results of the next section, we will give conditions under which this inverse exists for a sufficiently large number of observations.

3. Main results. In this section, we derive some asymptotic results for the MINQE(U, I) of the parameters of a covariance model for a Gaussian measure on a separable Hilbert space $\mathscr{H}$ without explicit reference to spatial covariance functions. At the end of this section, we will consider these results in the context of Gaussian random fields.

Consider $\{X_l\}_{l=1}^{\infty}$, a sequence of jointly Gaussian random variables forming a basis for $\mathscr{H}$. Suppose the true Gaussian measure on $\{X_l\}_{l=1}^{\infty}$ is given by (m, V), where m is the linear operator on $\mathscr{H}$ satisfying

$$m(X_l) = EX_l$$

and V is the bilinear operator on $\mathscr{H} \times \mathscr{H}$ satisfying

$$V(X_l, X_m) = \operatorname{cov}(X_l, X_m).$$

Suppose our (possibly incorrect) model for the mean of $\{X_l\}_{l=1}^{\infty}$ is

$$\text{(3.1)} \qquad EX_l = \beta' f_l,$$

where $f_1, f_2, \ldots$ are specified vectors. Our model for the covariance operator is

$$K_\theta(X_l, X_m) = \sum_{i=1}^{k} \theta_i K_i(X_l, X_m),$$

where $K_1, \ldots, K_k$ are specified bilinear operators such that $K_i(X_l, X_m)$ gives the covariance of X_l and X_m under the Gaussian measure $(0, K_i)$. We consider estimating θ using the MINQE(U, I) with starting value of θ denoted by α. Analogous to (2.2), we will assume

$$\text{(3.2)} \qquad K_\alpha \text{ is strictly positive definite,}$$

by which we mean that for all finite N and real constants $a_1, \dots, a_N$,

$$\sum_{i=1}^{N} a_i a_j K_\alpha(X_i, X_j) \geq 0,$$

with equality if and only if $a_1 = \cdots = a_N = 0$. Now let

$$r = \lim_{N \to \infty} \operatorname{rank}(f_1, \dots, f_N)$$

and

$$N_0 = \min_N \{\operatorname{rank}(f_1, \dots, f_N) = r\}.$$

We can define a sequence of observations $\{y_l\}_{l=1}^{\infty}$ such that y_l is a linear combination of $X_1, \dots, X_{l+r}$ for $l = 1, 2, \dots,$

$$Ey_l = 0 \quad \text{if (3.1) holds}$$

and

$$K_\alpha(y_l, y_m) = \delta_{lm}, \tag{3.3}$$

where $\delta_{lm} = 1$ if $l = m$ and 0 otherwise. We can then define the MINQE(U, I) of the X_l's in terms of the y_l's. Define $Y_N = (y_1, \dots, y_N)'$ and

$$\psi_i(l, m) = K_i(y_l, y_m).$$

Let $\Psi_i(N)$ be the $N \times N$ matrix with lmth element $\psi_i(l, m)$. For $N \geq N_0 - r$, if the MINQE(U, I) of θ based on $X_1, \dots, X_{N+r}$ exists, it is given by

$$\left[\operatorname{tr} \Psi_i(N) \Psi_j(N)\right]^{-1} \left[Y_N' \Psi_i(N) Y_N\right]. \tag{3.4}$$

To determine the properties of this statistic, only the distribution of $\{y_l\}_{l=1}^{\infty}$ is needed. We will use the subscript y to denote a Gaussian measure on $\{y_l\}_{l=1}^{\infty}$. Corresponding to the measure $(\beta' f, K_i)$ on $\{X_l\}_{l=1}^{\infty}$, by (3.1) we have the measure $(0, K_i)_y$ on $\{y_l\}_{l=1}^{\infty}$, and corresponding to the true measure (m, V) on $\{X_l\}_{l=1}^{\infty}$, we have the measure $(m, V)_y$ on $\{y_l\}_{l=1}^{\infty}$. Note that for events measurable with respect to $\{y_l\}_{l=1}^{\infty}$, the measures (m, V) and $(m, V)_y$ give identical probabilities. Finally, let

$$\begin{aligned} G_N &= \left[g_{ij}(N)\right] \\ &= \left[\operatorname{tr}\{(\Psi_i(N) - I)(\Psi_j(N) - I)\}\right] \\ &= \left[\sum_{l, m=1}^{N} (\psi_i(l, m) - \delta_{lm})(\psi_j(l, m) - \delta_{lm})\right]. \end{aligned}$$

The following is a strengthening of a theorem stated without proof by Stein (1987a).

THEOREM 1. *Suppose* (3.2) *holds,*

$$(0, K_1)_y \equiv \cdots \equiv (0, K_k)_y \equiv (0, K_\alpha)_y \tag{3.5}$$

and there exists a constant $c > 0$ such that

$$(m, V)_y \equiv (0, cK_\alpha)_y. \tag{3.6}$$

Then there exists a finite-valued G such that

(i) $\lim_{N \to \infty} G_N = G$.

Under the additional assumption

$$\begin{pmatrix} G & e \\ e' & 0 \end{pmatrix} \text{ is invertible}, \tag{3.7}$$

we also have:

(ii) *For all N sufficiently large, $\hat{\theta}_N(\alpha)$, the* MINQE(U, I) *of θ based on Y_N, exists.*

(iii) $(N/2)^{1/2}(e'\hat{\theta}_N(\alpha) - c)/c \to_{\mathscr{L}} N(0,1)$ *under* $(m, V)_y$.

(iv) $\hat{\theta}_N(\alpha) \to \hat{\theta}(\alpha)$ *in L^2 under $(m, V)_y$, where $\hat{\theta}(\alpha)$ is a finite, well-defined random vector.*

Before we prove this theorem, some comments about (3.7) are in order. This assumption is sufficient to guarantee that the MINQE(U, I) exists for all N sufficiently large. However, using the fact that the MINQE(U, I) exists if and only if (2.1) has a solution for all vectors p, it is easy to see that if the MINQE(U, I) exists for $N = N_0$, it also exists for all $N \geq N_0$. Thus, to obtain (ii), it is possible to make the perhaps more natural assumption that the MINQE(U, I) exists for some N. Where we really need (3.7) is to obtain (3.14), which allows us to compute the L^2 limit of $\hat{\theta}_N(\alpha)$. It is possible that the existence of the MINQE(U, I) for some N implies (3.7), but I have been unable to prove this.

PROOF OF THEOREM 1. We first prove (i). Using (2.20) of Ibragimov and Rozanov [(1978), page 81] and the fact that $(0, K_\alpha)_y \equiv (0, K_i)_y$, we obtain $g_{ii}(N)$ converges to a finite limit g_{ii} as $N \to \infty$. We have

$$g_{ij}(N) = \sum_{l=1}^{N} \left(g_{ij}(l) - g_{ij}(l-1)\right),$$

where $g_{ij}(0) = 0$. This series is absolutely summable as $N \to \infty$, since

$$\begin{aligned}\sum_{l=1}^{\infty} \left|g_{ij}(l) - g_{ij}(l-1)\right| &\leq \sum_{l,m=1}^{\infty} \left|\psi_i(l,m) - \delta_{lm}\right|\left|\psi_j(l,m) - \delta_{lm}\right| \\ &\leq \left\{g_{ii}g_{jj}\right\}^{1/2}\end{aligned}$$

by the Cauchy–Schwarz inequality. Thus, $g_{ij}(N)$ converges to a finite limit and (i) obtains.

Using (3.3), we have

$$G_N\alpha = 0 \quad \text{for all } N, \tag{3.8}$$

and hence,

$$G\alpha = 0. \tag{3.9}$$

Define the vector

$$v_N = \left[\operatorname{tr}(\Psi_i(N) - I)\right]. \tag{3.10}$$

We need the following lemma.

LEMMA 1. *If* $\sum_{n=1}^{\infty} a_n^2 < \infty$, *then* $\sum_{n=1}^{N} a_n = o(N^{1/2})$.

PROOF. It is sufficient to prove the lemma in the case of $a_n \geq 0$ for all n. If the sequence is rearranged so that it is monotonically decreasing (which is possible since $a_n \to 0$ as $n \to \infty$), then $\sum_{n=1}^{N} a_n$ is increased for all N, so it suffices to consider the case of $\{a_n\}_{n=1}^{\infty}$ nonnegative and decreasing. In this case, $a_n^2 = o(n^{-1})$ by Proposition 1, part ii, of Shorack and Wellner [(1986, page 864]. Thus, $a_n = o(n^{-1/2})$ and the lemma easily follows. □

Now,

$$\sum_{l=1}^{\infty} (\psi_i(l, l) - 1)^2 \leq \sum_{l, m=1}^{\infty} (\psi_i(l, m) - \delta_{lm})^2 = g_{ii} < \infty,$$

which, by Lemma 1, implies

$$\sum_{l=1}^{N} (\psi_i(l, l) - 1) = o(N^{1/2}).$$

So

$$v_N = o(N^{1/2}), \tag{3.11}$$

by which we mean that each element of v_N is $o(N^{1/2})$. Now (3.9) implies $|G| = 0$, so the lower right-hand element of the inverse of the matrix in (3.7) is zero. Thus,

$$\begin{pmatrix} G & e \\ e' & 0 \end{pmatrix}^{-1} = \begin{pmatrix} B & d \\ d' & 0 \end{pmatrix}$$

for some symmetric matrix B and vector d. We must have $Gd = 0$ and $e'd = 1$. But (3.7) and (3.9) imply that $\operatorname{rank}(G) = k - 1$, and it follows that $d = \alpha$. That is,

$$\begin{pmatrix} G & e \\ e' & 0 \end{pmatrix}^{-1} = \begin{pmatrix} B & \alpha \\ \alpha' & 0 \end{pmatrix}, \tag{3.12}$$

where $BG + \alpha e' = I$ and

$$Be = 0. \tag{3.13}$$

Now, to prove (ii), by (3.4) it is sufficient to show that

$$Q_N = \left[\operatorname{tr} \Psi_i(N)\Psi_j(N)\right]$$

is invertible for all N sufficiently large. We have

$$Q_N = G_N + ev'_N + v_N e' + Nee'.$$

By (i) and (3.7),

$$\begin{pmatrix} G_N & e \\ e' & 0 \end{pmatrix}$$

is invertible for all N sufficiently large. Applying (3.8),

$$\begin{pmatrix} G_N & e \\ e' & 0 \end{pmatrix}^{-1} = \begin{pmatrix} B_N & \alpha \\ \alpha' & 0 \end{pmatrix},$$

where

$$B_N \to B \tag{3.14}$$

and

$$B_N e = 0 \tag{3.15}$$

for N sufficiently large. By straightforward calculation, we can show

$$Q_N B_N = I + e(B_N v_N - \alpha)'$$

and

$$Q_N(B_N v_N - \alpha) = (v'_N B_N v_N - 2\alpha' v_N - N)e,$$

from which it follows that

$$Q_N^{-1} = B_N + (N + 2\alpha' v_N - v'_N B_N v_N)^{-1}(B_N v_N - \alpha)(B_N v_N - \alpha)' \tag{3.16}$$

if $N + 2\alpha' v_N - v'_N B_N v_N \neq 0$. But $N + 2\alpha' v_N - v'_N B_N v_N > 0$ for all N sufficiently large by (3.11) and (3.14).

To prove (iii) and (iv), we first consider the special case where $(m, V)_y = (0, cK_\alpha)_y$ for some positive c. Using (3.4), (3.15), (3.16) and $\alpha' e = 1$,

$$\begin{aligned} e'\hat{\theta}_N(\alpha) &= -(N + 2\alpha' v_N - v'_N B_N v_N)^{-1}(B_N v_N - \alpha)'[Y'_N \Psi_i(N) Y_N] \\ &= (N + 2\alpha' v_N - v'_N B_N v_N)^{-1}\{Y'_N Y_N - v'_N B_N [Y'_N(\Psi_i(N) - I)Y_N]\} \\ &\qquad\qquad \text{[using (3.3) and (3.15)]} \\ &= c + (N + 2\alpha' v_N - v'_N B_N v_N)^{-1}(Y'_N Y_N - cN) \\ &\quad - (N + 2\alpha' v_N - v'_N B_N v_N)^{-1}(v'_N B_N S_N + 2c\alpha' v_N), \end{aligned} \tag{3.17}$$

where

$$S_N = [s_i(N)] = \left[\sum_{l, m=1}^{N} (y_l y_m - c\delta_{lm})(\psi_i(l, m) - \delta_{lm})\right].$$

Now, $2\alpha' v_N - v'_N B_N v_N = o(N)$ and $Y'_N Y_N \sim c\chi^2_N$ under $(0, cK_\alpha)_y$, so

$$N^{1/2}(N + 2\alpha' v_N - v'_N B_N v_N)^{-1}(Y'_N Y_N - cN) \to_{\mathscr{L}} N(0, 2c^2) \quad \text{under } (0, cK_\alpha)_y. \tag{3.18}$$

Let $\mathscr{F}_N$ be the sigma field generated by the random vector Y_N. We can easily show that $\{s_i(N), \mathscr{F}_N\}_{N=1}^{\infty}$ is a martingale under $(0, cK_\alpha)_y$ and

$$Es_i(N)^2 = 2c^2 g_{ii}(N) \le 2c^2 g_{ii} < \infty,$$

so it is bounded in L^2, hence in L^1. Thus, by the martingale convergence theorem [Chung (1974), page 334] $s_i(N)$ converges almost surely under $(0, cK_\alpha)_y$. That is, we can say

$$S_N \overset{\text{a.e.}}{\to} S$$

$$= \left[\lim_{N\to\infty} \sum_{l,m=1}^{N} (y_l y_m - c\delta_{lm})(\psi_i(l,m) - \delta_{lm})\right] \quad \text{under } (0, cK_\alpha)_y. \tag{3.19}$$

Since $v_N' B_N v_N = o(N)$ and $v_N = o(N^{1/2})$, it follows that the last term in (3.17) is $o_p(N^{-1/2})$. Thus, using (3.17) and (3.18), we obtain (iii) when $(m, V)_y = (0, cK_\alpha)_y$. We also have

$$\hat{\theta}_N(\alpha) = (e'\hat{\theta}_N(\alpha))\alpha + B_N S_N + B_N v_N(c - e'\hat{\theta}_N(\alpha))$$
$$\overset{\text{pr}}{\to} c\alpha + BS \quad \text{under } (0, cK_\alpha)_y,$$

so that we have convergence in probability in (iv) when $(m, V)_y = (0, cK_\alpha)_y$. We see that an explicit expression for $\hat{\theta}(\alpha)$ is given by $c\alpha + BS$.

More generally, suppose (3.6) holds: $(0, cK_\alpha)_y \equiv (m, V)_y$. The fact that S is a well-defined finite random vector under $(0, cK_\alpha)_y$ implies that S is a well-defined finite random vector under $(m, V)_y$. Also, again using (3.6),

$$\|\hat{\theta}_N(\alpha) - c\alpha - BS\| \overset{\text{pr}}{\to} 0 \quad \text{under } (0, cK_\alpha)_y$$

implies

$$\|\hat{\theta}_N(\alpha) - c\alpha - BS\| \overset{\text{pr}}{\to} 0 \quad \text{under } (m, V)_y.$$

Thus, we have convergence in probability in (iv). Of course, the distribution of S depends on $(m, V)_y$. To obtain (iii), we note that the last term in (3.17) is $o_p(N^{-1/2})$ under $(m, V)_y$, since it is $o_p(N^{-1/2})$ under $(0, cK_\alpha)_y$. Thus, it suffices to show that

$$N^{-1/2}(Y_N'Y_N - cN) \to_{\mathscr{L}} N(0, 2c^2) \quad \text{under } (m, V)_y. \tag{3.20}$$

Choose H_N orthogonal so that under $(m, V)_y$,

$$\tilde{Y}_N = H_N Y_N \sim N(\mu_N, cW_N),$$

where W_N is diagonal. Thus,

$$(\tilde{Y}_N - \mu_N)' W_N^{-1} (\tilde{Y}_N - \mu_N) \sim c\chi_N^2 \quad \text{under } (m, V)_y, \tag{3.21}$$

$$N^{-1/2}\left[(\tilde{Y}_N - \mu_N)' W_N^{-1} (\tilde{Y}_N - \mu_N) - cN\right] \to_{\mathscr{L}} N(0, 2c^2) \quad \text{under } (m, V)_y \tag{3.22}$$

and, under $(m, V)_y$,

$$N^{-1/2} E\left\{\tilde{Y}_N'\tilde{Y}_N - (\tilde{Y} - \mu_N)' W_N^{-1} (\tilde{Y}_N - \mu_N)\right\}^2$$
$$= N^{-1/2}\left\{(\mu_N'\mu_N)^2 + 4c\mu_N' W_N \mu_N - 2c(\mu_N'\mu_N)\operatorname{tr}(I - W_N) + 2c^2 \operatorname{tr}(I - W_N)^2\right\}. \tag{3.23}$$

Now, (3.6) implies $\operatorname{tr}(I - W_N)^2$ and $\mu'_N\mu_N$ both are bounded as $N \to \infty$ [Ibragimov and Rozanov (1978), (2.9) on page 76] and it follows that $\mu'_N W_N \mu_N$ is bounded as $N \to \infty$. Since $\operatorname{tr}(I - W_N) = \operatorname{tr}(I - V_N)$, where cV_N is the actual covariance matrix of Y_N, and $\operatorname{tr}(I - V_N)^2 = \operatorname{tr}(I - W_N)^2$ is bounded, we have $\operatorname{tr}(I - W_N) = o(N^{1/2})$ by Lemma 1. Therefore, (3.23) tends to zero as $N \to \infty$. Combining this fact with (3.22), we obtain (3.20) and hence (iii).

To finish the proof of Theorem 1, we consider the L^2 convergence of $\hat{\theta}_N(\alpha)$. We first show that $S(N) \to S$ in L^2 under $(m, V)_y$. Since $S(N) \to S$ almost surely under $(m, V)_y$, it suffices to show that $S(N)$ is Cauchy in L^2 under $(m, V)_y$. We have for $N > N'$,

$$(s_i(N) - s_i(N'))^2 = \left\{ \sum_{l=N'+1}^{N} \sum_{m=N'+1}^{N} (y_l y_m - c\delta_{lm})(\psi_i(l, m) - \delta_{lm}) + 2 \sum_{m=N'+1}^{N} \sum_{l=1}^{N'} y_l y_m \psi_i(l, m) \right\}^2. \tag{3.24}$$

Let us define $\tau_l = Ey_l$ and $\psi(l, m) = \operatorname{cov}(y_l, y_m)$, where the lack of a subscript indicates that the expectations are taken under $(m, V)_y$. We have

$$\begin{aligned} E(y_l y_m y_q y_r) = {} & \psi(l, m)\psi(q, r) + \psi(l, r)\psi(m, q) \\ & + \psi(l, q)\psi(m, r) + \tau_q\tau_r\psi(l, m) \\ & + \tau_m\tau_r\psi(l, q) + \tau_m\tau_q\psi(l, r) \\ & + \tau_l\tau_r\psi(m, q) + \tau_l\tau_q\psi(m, r) + \tau_l\tau_m\psi(q, r). \end{aligned}$$

Thus, we can show

$$\begin{aligned} & E\left\{ \sum_{l=N'+1}^{N} \sum_{m=N'+1}^{N} (y_l y_m - c\delta_{lm})(\psi_i(l, m) - \delta_{lm}) \right\}^2 \\ & \quad = \left\{ \sum_{lm} (\psi_i(l, m) - \delta_{lm})(\psi(l, m) - c\delta_{lm}) \right\}^2 \\ & \qquad + \left\{ \sum_{lm} (\psi_i(l, m) - \delta_{lm})(\psi(l, m) - c\delta_{lm}) \right\} \left\{ \sum_{lm} \tau_l\tau_m(\psi_i(l, m) - \delta_{lm}) \right\} \\ & \qquad + 2 \sum_{lmqr} (\psi_i(l, m) - \delta_{lm})(\psi_i(q, r) - \delta_{qr})\psi(l, r)\psi(m, q) \\ & \qquad + 4 \sum_{lmqr} \tau_m\tau_q\psi(l, r)(\psi_i(l, m) - \delta_{lm})(\psi_i(q, r) - \delta_{qr}), \end{aligned} \tag{3.25}$$

where each index in the summations has limits $N' + 1$ and N. We have

$$\sum_{l, m=1}^{\infty} (\psi_i(l, m) - \delta_{lm})^2 = g_{ii} < \infty.$$

Using (3.6), we similarly have [see (2.20) of Ibragimov and Rozanov (1978), page 81]

$$\sum_{l,m=1}^{\infty} (\psi(l,m) - c\delta_{lm})^2 < \infty$$

and [see the proof of Theorem 3 of Ibragimov and Rozanov (1978), page 78]

$$\sum_{l=1}^{\infty} \tau_l^2 < \infty.$$

Thus, by applying the Cauchy–Schwarz inequality, we can easily show the first two terms in (3.25) tend to zero as $N, N' \to \infty$. By repeated application of the Cauchy–Schwarz inequality, we can obtain the general inequalities

$$\left|\sum_{lmq} a_{lm} b_{mq} c_{ql}\right| \le \left(\sum_{lm} a_{lm}^2\right)^{1/2} \left(\sum_{mq} b_{mq}^2\right)^{1/2} \left(\sum_{ql} c_{ql}^2\right)^{1/2} \tag{3.26}$$

and

$$\left|\sum_{lmqr} a_{lm} b_{mq} c_{qr} d_{rl}\right| \le \left(\sum_{lm} a_{lm}^2\right)^{1/2} \left(\sum_{mq} b_{mq}^2\right)^{1/2} \left(\sum_{qr} c_{qr}^2\right)^{1/2} \left(\sum_{rl} d_{rl}^2\right)^{1/2}. \tag{3.27}$$

Thus, considering the third term in (3.25), we have

$$\begin{aligned}
&\left|\sum_{lmqr} (\psi_i(l,m) - \delta_{lm})(\psi_i(q,r) - \delta_{qr})\psi(l,r)\psi(m,q)\right| \\
&\quad \le \sum_{lmqr} |\psi_i(l,m) - \delta_{lm}||\psi_i(q,r) - \delta_{qr}| \\
&\qquad \times\left(|\psi(l,r) - c\delta_{lr}| + c\delta_{lr}\right)\left(|\psi(m,q) - c\delta_{mq}| + c\delta_{mq}\right) \\
&\quad = \sum_{lmqr} |\psi_i(l,m) - \delta_{lm}||\psi_i(q,r) - \delta_{qr}||\psi(l,r) - c\delta_{lr}||\psi(m,q) - c\delta_{mq}| \\
&\qquad + 2c\sum_{lmq} |\psi_i(l,m) - \delta_{lm}||\psi_i(q,l) - \delta_{ql}||\psi(m,q) - c\delta_{mq}| \\
&\qquad + c^2\sum_{lm} (\psi_i(l,m) - \delta_{lm})^2.
\end{aligned}$$

Applying (3.26) and (3.27), we have that this expression tends to zero as $N, N' \to \infty$. We can also show the fourth term in (3.25) tends to zero as $N, N' \to \infty$, so we have that (3.25) tends to zero as $N, N' \to \infty$. Using similar manipulations, we can obtain

$$\lim_{N, N' \to \infty} E\left(\sum_{m=N'+1}^{N} \sum_{l=1}^{N'} y_l y_m \psi_i(l,m)\right)^2 = 0.$$

Thus, $E(s_i(N) - s_i(N'))^2$ tends to zero as $N, N' \to \infty$, so $s_i(N)$ is Cauchy in L^2 under $(m, V)_y$. Combining this result with (3.17), (3.21) and (3.23), we obtain $e'\hat{\theta}_N(\alpha) \to c$ in L^2 under $(m, V)_y$. The L^2 convergence of $\hat{\theta}_N(\alpha)$ follows, completing the proof of (iv) of Theorem 1. □

From (3.17), we also have

$$E(e'\hat{\theta}_N(\alpha)) = c + o(N^{-1/2}),\quad \operatorname{var}(e'\hat{\theta}_N(\alpha)) \sim 2c^2/N. \tag{3.28}$$

Of course, if $(m, V)_y = (0, K_\theta)_y$ for some k vector θ, then $E(e'\hat{\theta}_N(\alpha)) = e'\theta = c$ for all N.

To apply Theorem 1 to Gaussian random fields, let us suppose we model the mean and covariance functions as in (1.2) and (1.3) and consider a sequence of observations $z(x_1), z(x_2), \dots$ in R. Let us identify $\{z(x_l)\}_{l=1}^{\infty}$ with $\{X_l\}_{l=1}^{\infty}$ and make the obvious correspondences between the mean and covariance functions in Section 1 and the mean and covariance operators in Theorem 1. For example, the true mean function $m(x)$ defines a linear operator, which we will also call m, that satisfies

$$m(z(x)) = m(x),$$

and the true covariance function $V(x, x')$ defines a bilinear operator V that satisfies

$$V(z(x), z(x')) = V(x, x').$$

Now, if two Gaussian measures for the random field on R are equivalent, they are also equivalent on the (possibly smaller) Hilbert space generated by $\{z(x_l)\}_{l=1}^{\infty}$. Defining $\{y_l\}_{l=1}^{\infty}$ as in the beginning of this section in terms of $\{z(x_l)\}_{l=1}^{\infty}$, we further have that equivalence of Gaussian measures on the space generated by $\{z(x_l)\}_{l=1}^{\infty}$ implies equivalence on the space generated by $\{y_l\}_{l=1}^{\infty}$. Thus, compatibility of $K_1, \dots, K_k$ and K_α on R implies (3.5) and $(m, V) \equiv (\beta' f, cK_\alpha)$ on R for some β and $c > 0$ implies (3.6). Finally, (2.2) implies that $\{z(x_l)\}_{l=1}^{\infty}$ satisfies (3.2). Defining $\hat{\theta}_N(\alpha)$ and G_N as above, we have

COROLLARY 1. *For a Gaussian random field $z(\cdot)$ and a sequence of observations in R, if* (2.2) *and* (3.7) *hold,*

$$K_1, \dots, K_k \text{ and } K_\alpha \text{ are compatible on } R$$

and there exist $c > 0$ and β such that

$$(m, V) \equiv (\beta' f, cK_\alpha) \quad \text{on } R.$$

Then (i)–(iv) *of Theorem* 1 *hold.*

4. The MINQE(U, I) based on an infinite sequence. Since $\hat{\theta}(\alpha) = c\alpha + BS$ is the L^2 limit of a sequence of MINQE(U, I)'s, we might expect $\hat{\theta}(\alpha)$ to have a direct interpretation as a MINQE(U, I) of θ based on the sequence of Gaussian random variables $\{X_l\}_{l=1}^{\infty}$. Under the conditions given in Theorem 1, we now show that $\hat{\theta}(\alpha)$ does have such an interpretation.

Based on the sequence of contrasts $\{y_l\}_{l=1}^{\infty}$ as defined in the previous section, we wish to estimate $p'\theta$ for some fixed vector p. We will consider estimators of $p'\theta$ of the form

$$\lim_{N\to\infty} \sum_{l, m=1}^{N} u_{lm}(N) y_l y_m, \tag{4.1}$$

where the limit exists in L^2 under $(0, K_\alpha)_y$. We note that $p'\hat{\theta}(\alpha)$ is in this class. Generalizing the argument in the previous section, we can show that this limit exists in L^2 for any measure $(m, V)_y$ satisfying (3.6). Following the proof in Rao (1972), write

$$\lim_{N\to\infty} \sum_{l,m=1}^{N} u_{lm}(N) y_l y_m = p'\hat{\theta}(\alpha) + \lim_{N\to\infty} \sum_{l,m=1}^{N} v_{lm}(N) y_l y_m. \tag{4.2}$$

Now, $p'\hat{\theta}(\alpha)$ is the L^2 limit of unbiased estimators of $p'\theta$, so it is also unbiased. Thus, for the left-hand side of (4.2) to be an unbiased estimator of $p'\theta$, we must have

$$\lim_{N\to\infty} \sum_{l,m=1}^{N} v_{lm}(N)\psi_i(l, m) = 0 \quad \text{for } i = 1, \ldots, k. \tag{4.3}$$

We also have

$$\begin{aligned}
&\operatorname{cov}_\alpha\left(p'\hat{\theta}(\alpha), \lim_{N\to\infty} \sum_{l,m=1}^{N} v_{lm}(N) y_l y_m \right) \\
&= p'B\left(\operatorname{cov}_\alpha\left(s_i, \lim_{N\to\infty} \sum_{l,m=1}^{N} v_{lm}(N) y_l y_m \right)\right) \\
&= 2p'B\left(\lim_{N\to\infty} \sum_{l,m=1}^{N} v_{lm}(N)(\psi_i(l, m) - \delta_{lm})\right) = 0
\end{aligned}$$

using (4.3) and $Be = 0$, so that the two terms on the right-hand side of (4.2) are uncorrelated under $(0, K_\alpha)_y$. It follows that the variance under $(0, K_\alpha)_y$ of estimators of $p'\theta$ of the form given in (4.1) satisfying the unbiasedness conditions in (4.3) is minimized by $p'\hat{\theta}(\alpha)$. This estimator is unique in the sense that any other estimator of the form (4.1) that satisfies (4.3) and has the same variance as $p'\hat{\theta}(\alpha)$ under $(0, K_\alpha)_y$ has $p'\hat{\theta}(\alpha)$ as its L^2 limit under $(m, V)_y$.

5. Application to spatial covariance functions. Suppose we model a Gaussian process $z(\cdot)$ as in (1.1)–(1.3) with $f(\cdot)$ continuous on R compact and $K_\alpha(\cdot, \cdot)$ continuous on $R \times R$. We will show that $\hat{\theta}(\alpha)$, as defined in Section 3, is the same for any two dense sequence of points in R. That is, if $(m, V) \equiv (\beta' f, cK_\alpha)$ for some $c > 0$ and β, we get the same L^2 limit under (m, V) for $\hat{\theta}_N(\alpha)$ as $N \to \infty$ independent of the sequence of points, as long as it is dense in R.

Let C_0 be the class of finite contrasts under the model for the mean function given in (1.2): random variables of the form

$$\sum_{l=1}^{N} \lambda_l z(w_l)$$

satisfying

$$\sum_{l=1}^{N} \lambda_l f(w_l) = 0,$$

where $\lambda_1, \dots, \lambda_N$ are arbitrary, $w_1, \dots, w_N$ are points in R, and N ranges over the positive integers. Define C to be the L^2 closure of C_0 under $(0, K_\alpha)_y$, the Gaussian measure on the space of contrasts. By (2.6) of Ibragimov and Rozanov [(1978), page 76] C is also the closure under $(0, K_\theta)_y \equiv (0, cK_\alpha)_y$, $c > 0$. We need the following lemma.

LEMMA 2. *Suppose $\{x_l\}_{l=1}^\infty$ is a dense sequence of points in R, a compact set, $f(\cdot)$ is continuous on R and $K_\alpha(\cdot, \cdot)$ is continuous on $R \times R$. Then C is the L^2 closure under $(0, K_\alpha)_y$ of finite contrasts based on observations at $\{x_l\}_{l=1}^\infty$.*

PROOF. To prove the lemma, it suffices to show that if $X \in C$, X can be approximated arbitrarily well (in L^2) by a random variable of the form

$$\sum_{l=1}^N \lambda_l z(x_l)$$

satisfying $\sum \lambda_l f(x_l) = 0$. Since $X \in C$, given $\varepsilon > 0$, we can choose $w_1, \dots, w_J$ in R and $\lambda_1, \dots, \lambda_J$ satisfying $\sum \lambda_l f(w_l) = 0$ and

$$E_\alpha \left(X - \sum_{l=1}^J \lambda_l z(w_l) \right)^2 < \varepsilon, \tag{5.1}$$

where E_α indicates expectation under $(0, K_\alpha)_y$. Because $\{x_l\}_{l=1}^\infty$ is dense in R and $f(\cdot)$ and $K_\alpha(\cdot, \cdot)$ are continuous on R, we can choose $x_{a(1)}, \dots, x_{a(J)}$, a subset of $\{x_l\}_{l=1}^\infty$, satisfying

$$\left\| f(w_l) - f(x_{a(l)}) \right\| < \frac{\varepsilon}{J^{1/2} \|\lambda\|} \quad \text{for } l = 1, \dots, J, \tag{5.2}$$

where $\|\cdot\|$ denotes Euclidean norm and $\lambda = (\lambda_1, \dots, \lambda_J)'$, and

$$E_\alpha \left(\sum_{l=1}^J \lambda_l \big(z(w_l) - z(x_{a(l)}) \big) \right)^2 < \varepsilon. \tag{5.3}$$

Suppose $r = \sup \operatorname{rank}(f(w_1), \dots, f(w_q))$, where the supremum is taken over all $w_1, \dots, w_q \in R$ and q is the number of components in $f(\cdot)$. Since $f(\cdot)$ is continuous on R, there exist $x_{b(1)}, \dots, x_{b(r)} \in \{x_l\}_{l=1}^\infty$ such that $F_b = (f(x_{b(1)}), \dots, f(x_{b(r)}))$ has rank r. We will use these sites to adjust for the fact that $\sum \lambda_l z(x_{a(l)})$ is not quite a contrast. We have

$$\begin{aligned} \left\| \sum_{l=1}^J \lambda_l f(x_{a(l)}) \right\| &= \left\| \sum_{l=1}^J \lambda_l \big(f(x_{a(l)}) - f(w_l) \big) \right\| \\ &\le \|\lambda\| \left(\sum_{l=1}^J \left\| f(x_{a(l)}) - f(w_l) \right\|^2 \right)^{1/2} < \varepsilon, \end{aligned} \tag{5.4}$$

using (5.2). For $\eta = (\eta_1, \dots, \eta_r)'$, consider the equation

$$F_b \eta = \sum_{l=1}^J \lambda_l f(x_{a(l)}). \tag{5.5}$$

By the definition of F_b, F_b is of full rank and has the same rank as F, so the right-hand side of (5.5) is in the column space of F_b and it follows that (5.5) has a unique solution η. We also have $\|\eta\| < h\varepsilon$, where h is a constant depending only on F_b. (If we let F_b^* be an $r \times r$ nonsingular matrix made up of r rows from F_b, we can take h to be the matrix norm of F_b^{*-1}.) Then

$$\sum_{l=1}^{J} \lambda_l z(x_{a(l)}) - \sum_{l=1}^{r} \eta_l z(x_{b(l)})$$

is a contrast and

$$\begin{aligned}
E_\alpha&\left\{X - \left(\sum_{l=1}^{J} \lambda_l z(x_{a(l)}) - \sum_{l=1}^{r} \eta_l z(x_{b(l)})\right)\right\}^2 \\
&= E_\alpha\left\{\left(X - \sum_{l=1}^{J} \lambda_l z(w_l)\right) + \left(\sum_{l=1}^{J} \lambda_l\big(z(w_l) - z(x_{a(l)})\big)\right) + \sum_{l=1}^{r} \eta_l z(x_{b(l)})\right\}^2 \\
&\le 2\left\{E_\alpha\left(X - \sum_{l=1}^{J} \lambda_l z(w_l)\right)^2 + E_\alpha\left(\sum_{l=1}^{J} \lambda_l\big(z(w_l) - z(x_{a(l)})\big)\right)^2\right. \\
&\qquad\qquad \left. + E_\alpha\left(\sum_{l=1}^{r} \eta_l z(x_{b(l)})\right)^2\right\} \\
&\le 2\{\varepsilon + \varepsilon + Lr^2h^2\varepsilon^2\},
\end{aligned}$$

where

$$L = \sup_{w \in R} \operatorname{var}(z(w)).$$

Now ε and X are arbitrary, so we can conclude that C is contained in the L^2 closure of the contrasts of a finite number of the $z(x_i)$'s and since this L^2 closure is trivially contained in C, Lemma 2 is obtained. □

Under the conditions in Lemma 2, we have just shown that C is separable, since the contrasts of any dense but countable sequence of observations will serve as a basis for C. Suppose $\{y_l\}_{l=1}^{\infty}$ and $\{\tilde{y}_l\}_{l=1}^{\infty}$ are two orthonormal bases [under $(0, K_\alpha)_y$] for C. Define $a_{lq} = \operatorname{cov}_\alpha(y_l, \tilde{y}_q)$. We have

$$y_l = \sum_{q=1}^{\infty} a_{lq} \tilde{y}_q,$$

where the sum converges in L^2. Consider

$$\begin{aligned}
Q &= \lim_{N \to \infty} Q_N \\
&= \lim_{N \to \infty} \sum_{l,m=1}^{N} u_{lm}(N) y_l y_m,
\end{aligned}$$

where the limit exists in L^2. We have

$$Q_N = \lim_{J\to\infty} \sum_{l,m=1}^{N} u_{lm}(N) \sum_{q,r=1}^{N} u_{lm}(N) \sum_{q,r=1}^{J} a_{lq} a_{mp} \tilde{y}_q \tilde{y}_r$$
$$= \lim_{J\to\infty} \tilde{Q}_N(J),$$

where

$$\tilde{Q}_N(J) = \sum_{q,r=1}^{J} \left(\sum_{l,m=1}^{N} u_{lm}(N) a_{lq} a_{mr} \right) \tilde{y}_q \tilde{y}_r.$$

For a given N, we can choose J_N increasing in N such that

$$E_\alpha\big(Q_N - \tilde{Q}_N(J_N)\big)^2 < 1/N.$$

Since $Q_N \to Q$ in L^2, it follows that $\tilde{Q}_N(J_N) \to Q$ in L^2. Define $\tilde{Q}_N = \tilde{Q}_P(J_P)$ for $J_P \le N < J_{P+1}$, so $\tilde{Q}_N \to Q$ in L^2. Thus, the space of random variables defined by (4.1) is independent of the orthonormal basis chosen. By Lemma 2, any sequence of observations at a dense set of points in R can be used to form a basis for C, so the space of random variables of the form (4.1) is the same for any dense sequence of observations in R. We can conclude:

THEOREM 2. *Under the conditions of Corollary* 1 *and Lemma* 2, *the* MINQE(U, I) *of* $\hat{\theta}(\alpha)$ (*as defined in Theorem* 1) *is independent of the dense sequence of points* $\{x_l\}_{l=1}^{\infty}$ *in* R.

Thus, we are justified in calling $\hat{\theta}(\alpha)$ the MINQE(U, I) of θ based on observing $z(\cdot)$ everywhere in R.

6. Discussion. While the results of the previous sections provide important information about the asymptotic behavior of MINQE(U, I)'s of a spatial covariance function, there are many unresolved issues. In particular, in the expression $\hat{\theta} = c\alpha + BS$ for the MINQE(U, I) based on an infinite sequence of observations, we do not have an explicit expression for the distribution of the random vector S. It is not difficult to compute the mean and covariance matrix of S under the true Gaussian measure, but they do not tell the whole story since S, in general, will not be multivariate normal. In fact, under the conditions of Theorem 1, it seems unlikely that S will ever be multivariate normal except in the trivial case where θ has only one component. Another problem is that $K_{\hat{\theta}}(\cdot, \cdot)$ is not necessarily a positive definite function. Thus, in practice, some modification of the MINQE(U, I) is used [Stein (1987a)], and it is unclear what the properties of these modified estimators will be.

In the examples in Table 1 of Stein (1987a), the variances of MINQE(U, I)'s based on N observations in a compact region R showed remarkable insensitivity to the value of α, the starting choice for θ for both large and small N. Except for (3.28), which says that the variance of $e'\hat{\theta}_N(\alpha)$ is asymptotically independent of α, the results given here do not explain the insensitivity to α observed in these examples. These examples also suggest that N does not need to be very large for $\operatorname{var}(p'\hat{\theta}_N(\alpha))$ to be close to $\operatorname{var}(p'\hat{\theta}(\alpha))$ for p not proportional to e. The results in this paper do not give rates of convergence; however, if we were in a case where we could make a statement about v_N sharper than (3.11), it would be possible to give bounds on the errors. For example, if $v_N = O(1)$, then from (3.17) we can show for $(0, K_\theta)_y \equiv (0, cK_\alpha)_y$,

$$\operatorname{var}\big(e'\hat{\theta}_N(\alpha)\big) = 2c^2N^{-1} + O(N^{-2}) \quad \text{under } (0, K_\theta)_y,$$

which we can compare to (3.28). The behavior of v_N in the examples in Table 1 of Stein (1987a) is unknown. Considering the rapid convergence observed there, it would not be surprising if in fact $v_N = O(1)$ in at least some of those cases.

It is important to note the strong dependence of Theorem 1 on the assumption that the random field is Gaussian. As an example, suppose that $z(\cdot)$ is a stationary Gaussian process on $[0,1]$ with unknown mean μ and covariance function $K(x, x') = e^{-|x-x'|}$. We observe $t(x) = e^{z(x)}$, which is a stationary lognormal process with

$$Et(x) = e^{\mu+1/2} \equiv \theta^{1/2},$$
$$\operatorname{cov}(t(x), t(x')) = \theta\big(\exp\{e^{-|x-x'|}\} - 1\big).$$

Let $u(\cdot)$ be a Gaussian process on $[0,1]$ with the same first two moments as $t(x)$. Consider a sequence of distinct points in $[0,1]$, $\{x_l\}_{l=1}^{\infty}$, and let $x_{1N}, \dots, x_{NN}$ be $x_1, \dots, x_N$ rearranged in ascending order. We have that $\exp\{e^{-|x-x'|}\} - 1$ is compatible with $e(1 - |x - x'|)$ on $[0,1]$ [see Theorem 13, Chapter 3, of Ibragimov and Rozanov (1978)]. The MINQE(U, I) of θ based on $u(x_1), \dots, u(x_N)$ and assuming $\theta e(1 - |x - x'|)$ is the correct covariance function is

$$\sum_{i=2}^{N} \frac{\{u(x_{iN}) - u(x_{i-1,N})\}^2}{x_{iN} - x_{i-1,N}}. \tag{6.1}$$

Applying (3.28), we have that the asymptotic variance of this estimator is $2\theta^2e^2/N$. However, if we replace $u(\cdot)$ by $t(\cdot)$ in (6.1), we can show that the variance of this estimator does not tend to zero as N increases. Thus, this estimator is not consistent, at least not in L^2, despite the fact that $t(\cdot)$ has finite moments of every order. Of course, if we knew $t(\cdot)$ was lognormal, we could analyze the process on the log scale and avoid this problem. Cressie (1985) suggests a technique for choosing a pointwise transformation of a random field so that the transformed field is more nearly Gaussian. Unfortunately, a continuous random field need not be pointwise transformable to a Gaussian field, in which case, it is unclear how, or even if, one should estimate the covariance structure of the field.

Stein (1988) considers the asymptotic effect of using a covariance function which is compatible with the correct covariance function on kriging as the number of observations in a compact set R increases. We briefly summarize those results. Suppose $K_0(\cdot,\cdot)$ and $K_1(\cdot,\cdot)$ are two possible compatible covariance functions on $R \times R$. We want to predict some element of the Hilbert space given by the closure in L^2 under $(\beta'f, K_0)$ of random variables of the form given in (1.4). Define $e_i(N)$ to be the error of this kriging predictor based on $z(x_1), \dots, z(x_N)$ assuming the mean function is $\beta' f(\cdot)$ and the covariance function is $K_i(\cdot, \cdot)$. Let $V_i(\cdot)$ denote variance under $(\beta' f, K_i)$. Since the kriging predictor is unchanged when the covariance function is multiplied by a constant, we have the following trivial extension of Theorems 1 and 2 in Stein (1988).

THEOREM 3. *Suppose that for some $c > 0$, $K_0(\cdot, \cdot)$ and $cK_1(\cdot, \cdot)$ are compatible covariance functions on a compact set R and $\{x_l\}_{l=1}^{\infty}$ is a sequence of points in R. If*

$$\lim_{N \to \infty} V_0(e_0(N)) = 0, \tag{6.2}$$

then

$$\lim_{N \to \infty} \frac{V_0(e_0(N))}{V_0(e_1(N))} = 1 \tag{6.3}$$

and

$$\lim_{N \to \infty} \frac{V_0(e_1(N))}{V_1(e_1(N))} = c^{-1}. \tag{6.4}$$

Suppose $K_0(\cdot, \cdot)$ is the correct covariance function on $R \times R$. Then the variance of the kriging prediction error using $K_1(\cdot, \cdot)$ is asymptotically equivalent to the variance of the best linear unbiased predictor (6.3). Also, the ratio of the actual variance of the prediction error $e_1(N)$ to what we believe the variance of $e_1(N)$ is if we assume $K_1(\cdot, \cdot)$ is the correct covariance function tends to c as N increases (6.4). The assumption in (6.2) says that the best linear unbiased predictor is consistent, which will be true, for example, if $f(x)$ is continuous on R, $K_0(\cdot, \cdot)$ is continuous on $R \times R$, the linear functional being predicted is $z(x)$, and x is a limit point of $\{x_l\}_{l=1}^{\infty}$ [Stein (1987b)]. Theorem 3 does not apply if x is not a limit point of $\{x_l\}_{l=1}^{\infty}$. We note that Theorem 3 does not assume that $z(\cdot)$ itself is Gaussian, but only that the Gaussian measures on R with zero means and covariance functions $K_0(\cdot, \cdot)$ and $cK_1(\cdot, \cdot)$ are equivalent. Further results, including some bounds on the rates of convergence in (6.3) and (6.4), are given in Stein (1987b).

Theorem 3 says that to obtain asymptotically efficient kriging predictors (relative to the best linear unbiased predictor), it is sufficient to use a covariance function $K_1(\cdot, \cdot)$ such that $cK_1(\cdot, \cdot)$ is compatible with the true covariance function on R for some $c > 0$. To obtain asymptotically consistent values of the prediction error variance, then we also need to know the constant c. Relating this result to Theorem 1, we see that for a vector p not proportional to e, while we cannot estimate $p'\theta$ consistently, for purposes of linear prediction, we can get $p'\theta$ wrong and still get good predictions [in the sense that (6.3) and (6.4) hold], since as noted in Section 1, there exist θ and θ^* such that $p'\theta \neq p'\theta^*$ but K_θ and K_{θ^*} are compatible. Furthermore, from (6.4), $e'\theta$ will asymptotically control the ratio of the actual mean square prediction error to what we think the mean square prediction error is, and we can estimate $e'\theta$ consistently. Thus, for a

Gaussian random field, the one linear combination of θ that has a nonnegligible impact asymptotically on linear predictions is the one linear combination the MINQE(U, I) estimates consistently. Of course, since K_0 and K_1 in Theorem 3 are fixed covariance functions, this result does not directly apply to the problem of kriging with an estimated covariance function.

These results can be viewed as an example of what Dawid (1984) calls Jeffreys's law, which might be paraphrased as saying that aspects of a model which cannot be estimated consistently will, under appropriate conditions, have a negligible asymptotic impact on predictions. A mathematical embodiment of this law is given by a theorem due to Blackwell and Dubins (1962), which shows that, in great generality, equivalent prior distributions yield asymptotically indistinguishable conditional distributions. In fact, an alternative proof of Theorem 3, as well as some generalizations, can be obtained using their result. This connection will be further explored in future work.

Acknowledgment. The author would like to thank the Associate Editor, whose comments greatly improved the exposition of this paper.

REFERENCES

BLACKWELL, D. and DUBINS, L. E. (1962). Merging of opinions with increasing information. *Ann. Math. Statist.* **33** 882–886.

CHUNG, K. L. (1974). *A Course in Probability Theory*, 2nd ed. Academic, New York.

CRESSIE, N. (1985). Fitting variogram models by weighted least squares. *J. Internat. Assoc. Math. Geol.* **17** 563–586.

DAWID, A. P. (1984). Present position and potential developments: Some personal views. *J. Roy. Statist. Soc. Ser. A* **147** 278–292.

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

IBRAGIMOV, I. A. and ROZANOV, Y. A. (1978). *Gaussian Random Processes*. Springer, New York.

JOURNEL, A. G. and HUIJBREGTS, C. J. (1978). *Mining Geostatistics*. Academic, New York.

KITANIDIS, P. K. (1983). Statistical estimation of polynomial generalized covariance functions and hydrologic applications. *Water Resources Research* **19** 909–921.

KITANIDIS, P. K. (1985). Minimum-variance unbiased quadratic estimation of covariances of regionalized variables. *J. Internat. Assoc. Math. Geol.* **17** 195–208.

KUO, H. (1975). *Gaussian Measures in Banach Spaces. Lecture Notes in Math.* **463**. Springer, New York.

MARDIA, K. V. and MARSHALL, R. J. (1984). Maximum likelihood estimation of models for residual covariance in spatial regression. *Biometrika* **71** 135–146.

MARSHALL, R. J. and MARDIA, K. V. (1985). Minimum norm quadratic estimation of the components of spatial covariance. *J. Internat. Assoc. Math. Geol.* **17** 517–525.

MATHERON, G. (1973). The intrinsic random functions and their applications. *Adv. in Appl. Probab.* **5** 437–468.

RAO, C. R. (1971). Minimum variance quadratic unbiased estimation of variance components. *J. Multivariate Anal.* **1** 445–456.

RAO, C. R. (1972). Estimation of variance and covariance components in linear models. *J. Amer. Statist. Assoc.* **67** 112–115.

RAO, C. R. (1973). *Linear Statistical Inference and Its Applications*, 2nd ed. Wiley, New York.

RAO, C. R. (1979). MINQE theory and its relation to ML and MML estimation of variance components. *Sankhyā Ser. B* **41** 138–153.

SHORACK, G. and WELLNER, J. (1986). *Empirical Processes with Applications to Statistics*. Wiley, New York.

STEIN, M. L. (1987a). Minimum norm quadratic estimation of spatial variograms. *J. Amer. Statist. Assoc.* **82** 765–772.

STEIN, M. L. (1987b). Uniform asymptotic optimality of linear predictions of a random field using an incorrect second-order structure. Unpublished.

STEIN, M. L. (1988). Asymptotically efficient prediction of a random field with a misspecified covariance function. *Ann. Statist.* **16** 55–63.

YADRENKO, M. I. (1983). *Spectral Theory of Random Fields*. Optimization Software, New York.

DEPARTMENT OF STATISTICS
UNIVERSITY OF CHICAGO
5734 UNIVERSITY AVENUE
CHICAGO, ILLINOIS 60637