LOWER BOUNDS FOR MINIMUM COVARIANCE MATRICES IN TIME SERIES REGRESSION PROBLEMS[1]

By N. Donald Ylvisaker

University of Washington

1. Introduction. Linear regression problems in time series analysis have attracted considerable attention in the literature over the past decade and a half. Much of this work has been directed towards obtaining conditions under which least squares estimates have asymptotic efficiency 1 amongst the class of linear estimates, as for example, in [2], [3] and [5]. Asymptotic efficiency here refers to the limiting behavior of the covariance matrix of least squares estimates relative to the covariance matrix of minimum variance unbiased estimates as the observation set grows large, thus for the continuous parameter case, an observation set $[0, A]$ as $A \rightarrow \infty$. Over finite intervals, precise information about the relationship between these matrices is impossible to obtain unless one imposes a correlation structure of a very few special forms. Such questions beg answers as one does not usually know the best estimates or the minimum covariance matrix. Theorems of Parzen [8] give formal answers to these questions in terms of Gram matrices in appropriate reproducing kernel spaces. These theorems provide a uniform framework for such problems but do not usually give explicit solutions as they represent a restatement in terms of norms of functions. Since the framework is available, it is natural to inquire how the theory of these spaces may be used and we shall be concerned here with essentially one possibility of extending the norm structure for known examples to related classes of kernels. This will produce upper bounds on norms, which in turn will give lower bounds on minimum covariance matrices and lower bounds on efficiency. The classes of kernels for which these bounds are obtained are stationary and completely monotone or convex only. The latter case generalizes a result of Hájek [4] for the case of an unknown mean with a convex correlation function.

Section 2 is devoted to the basic inequality. Care is taken to state the result in some generality as we have not attempted to exhaust its potential here. In Section 3 we apply this to the above mentioned regression problems. Among the computed examples we show that the efficiency of the least squares estimate of the mean for an interval $[0, A]$ is at least $\frac{3}{4}$ if the correlation function is convex and is at least $\sim\frac{7}{8}$ if the correlation function is completely monotone, provided, in each case, the correlation function vanishes at infinity.

2. An inequality. Given a positive definite matrix R which is in the convex cone generated by the positive definite matrices K_α, and given a family of linearly independent vectors x_i, it is possible to bound above, in the sense of positive

Received 28 June 1963.

[1] This research was supported by the National Science Foundation under Grant G-25211.

definiteness, the Gram matrix $(x_i' R^{-1} x_j)$ in terms of the Gram matrices $(x_i' K_\alpha^{-1} x_j)$. In this section, we derive such an inequality for positive definite kernels which reduces to this statement for matrices. We view the inequality as one about Gram matrices in appropriate reproducing kernel spaces and use it later to produce bounds for minimum covariance matrices in certain regression problems, making use of the connection developed in this context by Parzen [8].

For T an abstract set, let $\mathcal{K}(T)$ be the class of real, finite valued, positive semi-definite kernels on $T \times T$. Corresponding to $K \varepsilon \mathcal{K}(T)$, there is a real Hilbert space of functions on T, denoted by $H(K)$ which is specified by the two conditions

(1) $K(\cdot, t) \varepsilon H(K)$ for all $t \varepsilon T$,

(2) if $f \varepsilon H(K)$, $(f, K(\cdot, t)) = f(t)$ for all $t \varepsilon T$.

The theory of these spaces is treated quite completely in [1]. In the class $\mathcal{K}(T)$ there is a partial ordering specified by $R \ll K$ if and only if $K - R \varepsilon \mathcal{K}(T)$, and we shall have occasion to make use of the way in which this partial ordering is reflected in the spaces $H(R)$ and $H(K)$.

To treat appropriate combinations of kernels generally, we let $(\Lambda, \mathcal{B}, \mu)$ be a probability space and suppose that $\{K_\lambda, \lambda \varepsilon \Lambda\}$ is a family of elements of $\mathcal{K}(T)$ having the property that $K\cdot(s, t)$ is measurable and in $L_1(d\mu)$ for each $(s, t) \varepsilon T \times T$. A straightforward computation shows that the kernel R on $T \times T$ given by $R = \int_\Lambda K_\lambda \, d\mu(\lambda)$ is also in $\mathcal{K}(T)$. We shall bound the Gram matrix in $H(R)$ of a family of functions by a suitable combination of the Gram matrices of the same family in $H(K_\lambda)$. To accomplish this, a lemma is given first to ease the proof. Note that henceforth all norms are indexed by the kernel in question.

Let $f_1, \cdots, f_n$ be a family of functions on T which are linearly independent. Let $\mathbf{f}$ be the vector valued function on T defined by $\mathbf{f}(t) = (f_1(t), \cdots, f_n(t))'$ and let $M(\mathbf{f}, K)$ denote the matrix $((f_i, f_j)_K)$.

LEMMA. *For $K \varepsilon \mathcal{K}(T)$, $M(\mathbf{f}, K) \ll M$ if and only if $F_M \ll K$ where $F_M(s, t) = \mathbf{f}(s)' M^{-1} \mathbf{f}(t)$ for $(s, t) \varepsilon T \times T$.*

PROOF. According to Theorems I and II on differences of kernels in [1], $F_M \ll K$ if and only if $H(F_M) \subset H(K)$ (set inclusion) and for every $g \varepsilon H(F_M)$, $\|g\|_{F_M}^2 \geqq \|g\|_K^2$. To see the structure of $H(F_M)$, let $m_1, \cdots, m_n$ be the columns of the positive definite matrix $M^{-\frac{1}{2}}$ so that $F_M(s, t) = \sum_{i=1}^n \varphi_i(s)\varphi_i(t)$ with $\varphi_i(\cdot) = \mathbf{f}(\cdot)' m_i$. In this form it is easily checked through conditions (1) and (2) that the functions $\varphi_1, \cdots, \varphi_n$ are a complete orthonormal basis for $H(F_M)$ and $\mathbf{f}(\cdot) = M^{\frac{1}{2}} M^{-\frac{1}{2}} \mathbf{f}(\cdot)$ implies $M(\mathbf{f}, F_M) = M$. If $F_M \ll K$ then $M(\mathbf{f}, K) \ll M(\mathbf{f}, F_M) = M$ and if $M(\mathbf{f}, K) \ll M$, $K \gg F_{M(\mathbf{f},K)} \gg F_M$.

For $R = \int_\Lambda K_\lambda \, d\mu(\lambda)$ we obtain an upper bound for $M(\mathbf{f}, R)$ in terms of $M(\mathbf{f}, K_\lambda)$ in the

THEOREM. *Suppose $M(\mathbf{f}, K_\lambda)^{-1}$ is measurable and in $L_1(d\mu)$, with $M(\mathbf{f}, K_\lambda)^{-1}$ taken to be the zero matrix if $f_i \not\varepsilon H(K_\lambda)$ for some i, then*

$$M(\mathbf{f}, R)^{-1} \gg \int_\Lambda M(\mathbf{f}, K_\lambda)^{-1} \, d\mu(\lambda).$$

PROOF. By the lemma above, $F_{M(\mathbf{f}, K_\lambda)} \ll K_\lambda$, and therefore

$$\int_\Lambda [K_\lambda - F_{M(\mathbf{f}, K_\lambda)}] \, d\mu(\lambda) = R - F_{M^*} \gg 0$$

where $M^{*-1} = \int_\Lambda M(\mathbf{f}, K_\lambda)^{-1} \, d\mu(\lambda)$. By reapplying the lemma one obtains $M(\mathbf{f}, R) \ll M^*$ or $M(\mathbf{f}, R)^{-1} \gg \int_\Lambda M(\mathbf{f}, K_\lambda)^{-1} \, d\mu(\lambda)$.

3. Applications. In this section, the inequality just derived will be applied to problems of linear regression on time series to obtain lower bounds on minimum covariance matrices with attending lower bounds on efficiency. Due to the dependence of the bound on information relative to the spaces $H(K_\lambda)$, we will at present be able to give explicit bounds when the underlying correlation structure is stationary and completely monotone or only convex.

Suppose one is allowed to observe $Y(t) = \sum_{i=1}^n \beta_i f_i(t) + X(t)$, $t \in T$, a subset of the real line, with $EX(t) \equiv 0$, $EX(s)X(t) = \rho^*(s - t)$, $\rho^*(0) = 1$. For minimum variance linear unbiased estimation of the β_i, we appeal to the following result due to Parzen [8],

THEOREM. *The covariance matrix of best unbiased estimators of the β_i is given by* $M(\mathbf{f}, \rho^*)^{-1}$.

Thus for a correlation function ρ^* appropriately spanned by correlation functions ρ_λ, the inequality of Section 2 is directly translated into a lower bound on the covariance matrix of best estimators.

Since one is dependent on knowledge of the $H(\rho_\lambda)$, attention is focussed on the classes defined below. For ρ a correlation function on the real line, $\rho(0) = 1$, let ρ_λ be defined by $\rho_\lambda(t) = \rho(\lambda t)$ and let $\mathcal{C}_\rho$ be a class of correlation functions on the real line given by

$$\mathcal{C}_\rho = \{\rho^* \mid \rho^* = \int_0^\infty \rho_\lambda \, dF(\lambda),\ F \text{ a distribution function on } (0, \infty)\}.$$

If $\rho(t) = \cos t$, then $\mathcal{C}_\rho$ is the class of all (real and normalized) correlations functions whose spectral distribution functions do not have a discontinuity at the origin. If $\rho(t) = (1 - |t|)^n$, $|t| \leqq 1$, then $\mathcal{C}_\rho$ is the class Γ_n, with $\rho^*(\infty) = 0$, defined in [5] and generalizing the convex or Pólya correlation functions. If $\rho(t) = e^{-|t|}$, then $\mathcal{C}_\rho$ is the class of all completely monotone correlation function with $\rho^*(\infty) = 0$. The classes $\mathcal{C}_\rho$ shrink in the following sense: if $\rho^* \in \mathcal{C}_\rho$ then $\mathcal{C}_{\rho^*} \subset \mathcal{C}_\rho$, for if

$$\rho^* = \int_0^\infty \rho_\lambda \, dF(\lambda), \qquad \bar{\rho} = \int_0^\infty \rho_\theta^* \, dG(\theta),$$

then

$$\bar{\rho} = \int_0^\infty \rho_\xi \, dH(\xi) \quad \text{where} \quad H(\xi) = \int_0^\infty F(\xi/\theta) \, dG(\theta).$$

The inequality then has the property

$$M(\mathbf{f}, \bar{\rho})^{-1} \gg \int_0^\infty M(\mathbf{f}, \rho_\theta^*)^{-1} \, dG(\theta) \gg \int_0^\infty \int_0^\infty M(\mathbf{f}, \rho_{\theta\lambda})^{-1} \, dF(\lambda) \, dG(\theta)$$
$$= \int_0^\infty M(\mathbf{f}, \rho_\xi)^{-1} \, dH(\xi),$$

i.e., it is increasingly good as the class decreases. At one extreme $\rho(t) = \cos(t)$, and in this case $f \in H(\rho_\lambda)$ for at most one choice of λ and so for most ρ^*, $\mathbf{f}$, the uninteresting result $M(\mathbf{f}, \rho^*)^{-1} \gg 0$. At the opposite extreme $\rho^* = \int_0^\infty \rho_\lambda^* \, dF(\lambda)$ where F has a unit jump at $\lambda = 1$, and there is actual equality.

For $\rho(t) = e^{-|t|}$ and $\rho(t) = 1 - |t|$, $|t| \leqq 1$, explicit bounds arise, since the structure of $H(\rho_\lambda)$ is known. Before exhibiting these, we note the isomorphisms due to translation and scale change, viz. $\|f\|_\rho^2$ on T is $\|f^b\|_\rho^2$ on T^{-b} with $f^b(t) = f(t + b)$ and T^{-b} the set T translated to the left by b, $\|f\|_\rho^2$ on T is $\|f_\lambda\|^2 \rho_\lambda$ on $T_{\lambda^{-1}}$ with $f_\lambda(t) = f(t\lambda)$ and $T_{\lambda^{-1}}$ the set T scaled by λ^{-1}. We shall take T to be either an interval or a finite collection of points and it is sufficient therefore to give the norm for $H(\rho)$ with $T = [0, A]$ or $T = \{t_1, \cdots, t_k\}$, $0 = t_1 < \cdots < t_k$. For $\rho(t) = e^{-|t|}$,

(a) $T = [0, A]$, $\|f\|_\rho^2 = \frac{1}{2} \int_0^A (f' + f)^2 + f^2(0)$, f absolutely continuous,

(b) $T = \{t_1, \cdots, t_k\}$, $\|f\|_\rho^2 = \sum_{i=1}^{k-1} \dfrac{(f(t_i)e^{t_{i+1}} - f(t_{i+1})e^{t_i})^2}{e^{2t_{i+1}} - e^{2t_i}} + f^2(t_k)$.

Case (a) is found in [8], while (b), which follows from (a), can also be obtained by noting that $\|f\|_\rho^2 = (f(t_1), \cdots, f(t_k))R^{-1}(f(t_1), \cdots, f(t_k))'$ with $R = (\rho(t_i - t_j))$. For $\rho(t) = 1 - |t|$, $|t| \leqq 1$, the norm structure is newly found and somewhat more tedious.

(a) $T = [0, A]$, A an integer

$$\|f\|_\rho^2 = [\textstyle\sum_{j=0}^{A} f(j)]^2/(A + 1)$$
$$+ [1/(A + 1)] \int_0^1 \textstyle\sum_{j=0}^{A-1} (j + 1)(A - j)f'^2(u + j)$$
$$+ [2/(A + 1)] \int_0^1 \textstyle\sum_{0 \leqq i < j < A} (n - j)(i + 1)f'(u + i)f'(u + j),$$

f absolutely continuous.

$T = [0, A]$, $A = (n - 1) + \delta$, $0 < \delta < 1$, define for an absolutely continuous function f, a continuous function f_c on $[0, n]$ by $f_c(t) = f(t)$, $t \in [0, A]$ and

$$f_c'(t + (n - 1)) = -(1/n) \sum_{j=0}^{n-2} (j + 1)f'(t + j) + c. \qquad \delta < t \leqq 1.$$

Then, $\|f\|_\rho^2 = \min_c \|f_c\|_\rho^2$, where $\|f_c\|_\rho^2$ is calculated from the previous case. The first part of this assertion is verified by checking the two conditions which specify $H(\rho)$. The second part follows from the theorem on restricted kernels in [1] and gives

$$\|f\|_\rho^2 = \min_{g=f \text{ on } [0,A]} \|g\|_\rho^2$$

where g is a function on $[0, n]$. That this is equivalent to the minimization given is easily checked, and we omit these unattractive details.

(b) $T = \{t_1, \cdots, t_k\}$. If $f(t) = \sum_{j=1}^{k} c_j \rho(t - t_j)$ for $t = t_i$, $i = 1, \cdots, k$,

then $\|f\|_\rho^2 = \sum_{i=1}^{k} \sum_{j=1}^{k} c_i c_j \rho(t_i - t_j)$.

No particular benefit accrues here from knowing $\| \quad \|_\rho^2$ on intervals. The solution of the equations above can be handled with knowledge of the spacing of the points t_i. One might also use the restricted kernel theorem as above, although it does not make the problem substantially easier.

We now take some easily computable, one parameter examples to illustrate the foregoing facts.

EXAMPLE 1. With $T = [0, A]$, $EY(t) = \beta$, $E(Y(s) - \beta)(Y(t) - \beta) = \rho^*(s - t) = \int_0^\infty e^{-\lambda|s-t|}\, dF(\lambda)$, note that $\|1\|_{\rho_\lambda}^2 = 1 + \lambda A/2$ and so minimum variance under $\rho^* \geqq \int_0^\infty [2/(2 + \lambda A)]\, dF(\lambda) = (2/A) \int_0^\infty e^{-2t/A} \rho^*(t)\, dt$.

EXAMPLE 2. In the same circumstances with $EY(t) = \beta f(t)$, minimum variance under ρ^*

$$\geqq \int_0^\infty \{2\lambda/[\int_0^A f'^2 + \lambda(f^2(0) + f^2(A)) + \lambda^2 \int_0^A f^2]\}\, dF(\lambda)$$
$$= (2/\int_0^A f^2) \int_0^\infty e^{-at}[\cos bt - a/b \sin bt]\rho^*(t)\, dt$$

where $a = [f^2(0) + f^2(A)]/2 \int_0^A f^2$ and $b = (\int_0^A f'^2 / \int_0^A f^2 - a^2)^{\frac{1}{2}}$.

Before doing similar computations with $\rho(t) = 1 - |t|$, $|t| \leqq 1$, let us note that we are writing a convex correlation function ρ^* as

$$\rho^* = \int_0^\infty \rho_\lambda\, dF(\lambda) = \int_0^{A^{-1}} \rho_\lambda\, dF(\lambda) + \int_{A^{-1}}^\infty \rho_\lambda\, dF(\lambda)$$
$$= F(A^{-1})\rho_1 + (1 - F(A^{-1}))\rho_2 .$$

For $T = [0, A]$, the initial inequality may be written as

$$M(\mathbf{f}, \rho^*)^{-1} \gg \int_0^\infty M(\mathbf{f}, \rho_\lambda)^{-1}\, dF(\lambda)$$
$$= \int_0^{A^{-1}} M(\mathbf{f}, \rho_\lambda)^{-1}\, dF(\lambda) + \int_{A^{-1}}^\infty M(\mathbf{f}, \rho_\lambda)^{-1}\, dp(\lambda).$$

Now $\rho_\lambda(t) = 1 - \lambda t$ on $[0, A]$ if $A \leqq \lambda^{-1}$ or $\lambda \leqq A^{-1}$, and therefore

$$\rho_1(t) = [1/F(A^{-1})] \int_0^{A^{-1}} (1 - \lambda t)\, dF(\lambda) = 1 - ([1/F(A^{-1})] \int_0^{A^{-1}} \lambda\, dF(\lambda))t, \quad t \in [0, A]$$

Since the norm structure is completely known in $H(\rho_1)$, another application of the basic inequality yields

$$M(\mathbf{f}, \rho^*)^{-1} \gg F(A^{-1})M(\mathbf{f}, \rho_1)^{-1} + (1 - F(A^{-1}))M(\mathbf{f}, \rho_2)^{-1}$$
$$\gg F(A^{-1})M(\mathbf{f}, \rho_1)^{-1} + \int_{A^{-1}}^\infty M(\mathbf{f}, \rho_\lambda)^{-1}\, dF(\lambda)$$

which represents an improvement for most $\mathbf{f}$.

EXAMPLE 3. With $T = [0, A]$, $EY(t) \equiv \beta$, $E(Y(s) - \beta)(Y(t) - \beta) = \rho^*(s - t) = \int_0^\infty \rho_\lambda(s - t)\, dF(\lambda)$, $\rho_\lambda(t) = 1 - \lambda|t|$, $|t| \leqq \lambda^{-1}$, $\|1\|_{\rho_\lambda}^2 = n(n + 1)/(2n - \lambda A)$ where $n = [\lambda A + 1]$. The inequality

$$\text{minimum variance under} \quad \rho^* \geqq \int_0^\infty (2n - A\lambda)/n(n + 1)\, dF(\lambda)$$

is that of Hájek [4] and the improvement above evaporates due to the nature of the regression function.

We turn then to some remarks about efficiency. It may be noticed that for certain regression functions, the kernel ρ has a least favorable character in the class $\mathcal{C}_\rho$ as regards efficiency of least squares estimators. Specifically, let $e_\rho(f, T)$ be the ratio of the minimum variance with regression function f to the variance of the least squares estimator. Then $e_\rho(f, T) = e_{\rho_\lambda}(f_{\lambda^{-1}}, T_\lambda)$ because of the scale isomorphism. Consider, for example, the case of an unknown mean with $T = [0, A]$,

$$e_\rho(1, [0, A]) = (\|1\|_\rho^2(2/A) \int_0^A (1 - u/A)\rho(u)\, du)^{-1}.$$

Suppose $e_\rho(1, [0, A])$ has a minimum value, in A, of e_0, and let $\rho^* = \int_0^A \rho_\lambda\, dF(\lambda)$.

$$e_{\rho^*}(1, [0, A]) \geqq \frac{\int_0^\infty \|1\|_{\rho_\lambda}^{-2}\, dF(\lambda)}{(2/A) \int_0^A (1 - u/A)\rho^*(u)\, du}$$

$$= \frac{\int_0^\infty \|1\|_{\rho_\lambda}^{-2}\, dF(\lambda)}{\int_0^\infty [(2/A) \int_0^A (1 - u/A)\rho_\lambda(u)\, du]\, dF(\lambda)} \geqq e_0,$$

as the ratio of the integrands is $e_{\rho_\lambda}(1, [0, A]) = e_\rho(1, [0, A\lambda^{-1}])$. Choosing $\rho(t) = e^{-|t|}$ this minimum is the minimum over A of $A^2/(2 + A)(A + e^{-A} + 1)$ and for $\rho(t) = 1 - |t|, |t| \leqq 1$, it is

$$\min\left(\min_{0 \leqq A \leqq 1} \frac{2 - A}{2} \frac{3}{3 - A}, \min_{A \geqq 1} \frac{2[A + 1] - A}{[A + 1][A + 2]} \frac{3A^2}{3A - 1}\right).$$

These minimums which then are lower bounds for the efficiency of the least squares estimator of the mean on intervals $[0, A]$ for completely monotone and convex correlation functions, are respectively, $\sim\frac{7}{8}$ and $\frac{3}{4}$. If, instead of taking $f \equiv 1$ so that $f_\lambda = f$, we consider f for which $f_\lambda = c(\lambda)f$, the same type of result will follow, as then

$$e_{\rho_\lambda}(f, T) = e_\rho(f_\lambda, T_{\lambda^{-1}}) = e_\rho(c(\lambda)f, T_{\lambda^{-1}}) = e_\rho(f, T_{\lambda^{-1}}).$$

With an appropriate definition of efficiency in the many parameter case, these results will hold for polynomial regression. For a regression of the form $\sum_{i=1}^n \beta_i f_i$, define the efficiency (cf. [6]) of the vector estimator $\bar{\beta}$ with respect to the best estimator $\hat{\beta}$ by

$$e_{\rho^*}(\mathbf{f}, [0, A]) = E\int_0^A [\hat{\beta}'\mathbf{f}(t) - \beta'\mathbf{f}(t)]^2\, dt / E\int_0^A [\bar{\beta}'\mathbf{f}(t) - \beta'\mathbf{f}(t)]^2\, dt$$

$$= \int_0^A \mathbf{f}(t)'M(\mathbf{f}, \rho^*)^{-1}\mathbf{f}(t)\, dt / \int_0^A \mathbf{f}(t)'Q_{\rho^*}(\mathbf{f})\mathbf{f}(t)\, dt$$

where $Q_{\rho^*}(\mathbf{f})$ is the covariance matrix of the estimator $\bar{\beta}$. If $\rho^* = \int_0^\infty \rho_\lambda\, dF(\lambda)$ and $\bar{\beta}$ is the vector of least squares estimates, then $Q_{\rho^*}(\mathbf{f}) = \int_0^\infty Q_{\rho_\lambda}(\mathbf{f})\, dF(\lambda)$ and

$$e_{\rho^*}(\mathbf{f}, [0, A]) \geqq \frac{\int_0^\infty [\int_0^A \mathbf{f}(t)'M(\mathbf{f}, \rho_\lambda)^{-1}\, \mathbf{f}(t)\, dt]\, dF(\lambda)}{\int_0^\infty [\int_0^A \mathbf{f}(t)'Q_{\rho_\lambda}(\mathbf{f})\mathbf{f}(t)\, dt]\, dF(\lambda)}.$$

The ratio of the integrands is again $e_{\rho_\lambda}(\mathbf{f}, [0, A])$. Changing scale, the right hand side becomes

$$\frac{\int_0^\infty [\int_0^{\lambda A} \mathbf{f}_{\lambda^{-1}}(t)' M(\mathbf{f}_{\lambda^{-1}}, \rho)^{-1} \mathbf{f}_{\lambda^{-1}}(t)\, dt]\, dF(\lambda)}{\int_0^\infty [\int_0^{\lambda A} \mathbf{f}_{\lambda^{-1}}(t)' Q_\rho(\mathbf{f}_{\lambda^{-1}}) \mathbf{f}_{\lambda^{-1}}(t)\, dt]\, dF(\lambda)},$$

with $M(\mathbf{f}_{\lambda^{-1}}, \rho)$ and $Q_\rho(\mathbf{f}_{\lambda^{-1}})$ calculated on $[0, A\lambda]$. Imposing the restriction that for each f_i, $f_i(\lambda^{-1}t) = c_i(\lambda^{-1}) f_i(t)$ for all $t \varepsilon [0, \lambda A]$,

$$e_{\rho^*}(\mathbf{f}, [0, A]) \geqq \frac{\int_0^\infty [\int_0^{\lambda A} \mathbf{f}(t)' M(\mathbf{f}, \rho)^{-1} \mathbf{f}(t)\, dt]\, dF(\lambda)}{\int_0^\infty [\int_0^{\lambda A} \mathbf{f}(t)' Q_\rho(\mathbf{f}) \mathbf{f}(t)\, dt]\, dF(\lambda)}$$

and the ratio of the integrands is $e_\rho(\mathbf{f}, [0, \lambda A])$.

4. Acknowledgment. I am grateful to Ingram Olkin for an alternative proof of the inequality of Section 2 stated specifically for positive definite matrices. He has pointed out that the inequality is apparently new in this context but can be handled quite easily by methods similar to those in [7].

REFERENCES

[1] Aronszajn, N. (1950). Theory of reproducing kernels. *Trans. Amer. Math. Soc.* **68** 337–404.

[2] Grenander, Ulf (1954). On the estimation of regression coefficients in the case of an autocorrelated disturbance. *Ann. Math. Statist.* **25** 252–272.

[3] Grenander, Ulf and Rosenblatt, Murray (1957). *Statistical Analysis of Stationary Time Series*. Wiley, New York.

[4] Hájek, J. (1956). Linear estimation of the mean value of a stationary random process with convex correlation function. *Čz. Math. J.* **6** 94–117. (Also in *Selected Translations in Math. Statist. and Prob.* **2** (1962) 41–61. Amer. Math. Soc., Providence.)

[5] Lévy, Paul (1962). Extensions dún théorème de D. Dugué et M. Girault. *Z. Wahrscheinlichkeitstheorie* **1** 159–173.

[6] Mann, Henry B. and Moranda, Paul B. (1954). On the efficiency of the least squares estimates of parameters in the Ornstein-Uhlenbeck process. *Sankhyā* **13** 351–358.

[7] Olkin, Ingram and Pratt, John W. (1958). A multivariate Tchebycheff inequality. *Ann. Math. Statist.* **29** 226–234.

[8] Parzen, Emanuel (1961). An approach to time series analysis. *Ann. Math. Statist.* **32** 951–989.