Stochastic Orders and Decision under Risk
IMS Lecture Notes – Monograph Series (1991)

ORDERING REGRESSION MODELS OF GAUSSIAN PROCESSES

BY HARALD LUSCHGY*

Universität Münster

We study information orderings of regression models of continuous time Gaussian processes. This is done in the framework of "comparison of experiments." The theory includes parametric, nonparametric and semiparametric models. Particular simple criteria for information orderings are obtained for linear models. Applications include the design problem for regression models with correlated Gaussian errors and the information contained in additional observation periods. Several examples are discussed.

1. Introduction. Informational inequalities between models of continuous time Gaussian processes are investigated. Let

$$X(t) = S(\theta, t) + Z(t), \quad t \in I, \tag{1.1}$$

where I is a compact subset of $\mathbb{R}$, Z is a zero-mean Gaussian process with covariance function K, θ is an unknown parameter belonging to an arbitrary parameter set Θ, and $S : \Theta \times I \to \mathbb{R}$ is a deterministic function. K and S are assumed to be known. We are interested in comparing the information contained in such models concerning inference about θ when Θ is fixed while S, K and I may vary from model to model. The basic ordering which will be considered is the same as the stochastic ordering of likelihood processes defined by convex functions.

The above model is an example of a signal-plus-noise model with deterministic signal S and noise process Z. Linear models of type (1.1) are given, for instance, by parametric regression models

$$X(t) = \sum_{i=1}^{r} \theta_i h_i(t) + Z(t),$$

---

* Research supported by a Heisenberg grant of the Deutsche Forschungsgemeinschaft.

AMS 1980 Subject Classification: Primary 62B15; 60G15

Regression models, Gaussian processes, information orderings, likelihood processes, deficiency.

where $\theta \in \mathbb{R}^r$ and h_i are known regression functions, nonparametric regression models

$$X(t) = \theta(t) + Z(t),$$

where θ is only assumed to belong to a (possibly infinite dimensional) given space Θ of regression functions and semiparametric regression models

$$X(t) = \sum_{i=1}^{r} \beta_i h_i(t) + \gamma(t) + Z(t),$$

where $\theta = (\beta, \gamma), \beta \in \mathbb{R}^r$ and γ belongs to some prescribed space of functions. For an example of a nonlinear model, consider the observation of a harmonic signal corrupted by additive noise

$$X(t) = \sin \theta t + Z(t),$$

where the frequency θ is unknown.

The systematic study of the comparison of nonparametric regression models for stochastic processes when no prior information about the regression function is available was started by Luschgy [15], [16].

Let us now briefly describe some basic concepts for comparing statistical experiments. For this topic we refer to Heyer [4], LeCam [13], Strasser [23] and Torgersen [26], [28]. Let $E = (\mathcal{X}, \mathcal{B}(\mathcal{X}), (P_\theta : \theta \in \Theta))$ and $F = (\mathcal{Y}, \mathcal{B}(\mathcal{Y}), (Q_\theta : \theta \in \Theta))$ be two experiments with the same parameter set Θ. Here $(P_\theta : \theta \in \Theta)$ and $(Q_\theta : \theta \in \Theta)$ are parameterized families of probability measures on the sample spaces $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$ and $(\mathcal{Y}, \mathcal{B}(\mathcal{Y}))$, respectively. Below we will restrict ourselves to homogeneous experiments with Polish sample spaces equipped with their Borel σ-algebras.

Definition 1. E is said to be at least as informative as F, $E \geq F$, if there is a Markov kernel M from $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$ to $(\mathcal{Y}, \mathcal{B}(\mathcal{Y}))$ such that

$$MP_\theta = Q_\theta \text{ for every } \theta \in \Theta, \tag{1.2}$$

where

$$MP_\theta(C) = \int M(x, C) dP_\theta(x), \quad C \in \mathcal{B}(\mathcal{Y}).$$

E and F are said to be equally informative, $E \sim F$, if $E \geq F$ and $F \geq E$.

It turns out that this ordering may be phrased in terms of finite subexperiments. For $\Gamma \subset \Theta$, let $E_\Gamma = (\mathcal{X}, \mathcal{B}(\mathcal{X}), (P_\theta : \theta \in \Gamma))$. Then $E \geq F$ if and only if $E_\Gamma \geq F_\Gamma$ for every finite subset Γ of Θ. Moreover, for $\Gamma = \{\theta_0, \theta_1, \cdots, \theta_n\} \subset \Theta$, $E_\Gamma \geq F_\Gamma$ holds if and only if

$$\int \varphi d\mathcal{L}(dP_{\theta_1}/dP_{\theta_o}, \cdots, dP_{\theta_n}/dP_{\theta_o} \mid P_{\theta_o}) \tag{1.3}$$

$$\geq \int \varphi d\mathcal{L}(dQ_{\theta_1}/dQ_{\theta_o},\cdots,dQ_{\theta_n}/dQ_{\theta_o} \mid Q_{\theta_o})$$

for every convex function $\varphi : (0,\infty)^n \to \mathbb{R}$.

This important criterion follows from a reduction by sufficiency and [28], Theorem 4.1. Note that the covariance of the likelihood process is increased by increasing the information.

The decision theoretic interpretation is based on the comparison of risk functions available in the two experiments. We have $E \geq F$ if and only if for every finite decision space D, for every bounded loss function $\ell : \Theta \times D \to \mathbb{R}$ and for every decision rule ρ in F there exists a decision rule σ in E such that

$$\begin{aligned}&\int_{\mathcal{X}}\int_{\mathcal{D}} \ell(\theta,a)\sigma(x,da)dP_\theta(x)\\ &\qquad \leq \int_{\mathcal{Y}}\int_{\mathcal{D}} \ell(\theta,a)\rho(y,da)dQ_\theta(y) \text{ for every } \theta\in\Theta.\end{aligned} \tag{1.4}$$

A related (weaker) ordering is the pointwise ordering of Hellinger transforms. The Hellinger transform H_E of E is defined on the set of prior distributions λ on Θ with finite support. If $\mathrm{supp}(\lambda) = \{\theta_1,\cdots,\theta_n\}$ and $\lambda(\theta_i) = \lambda_i$, then

$$H_E(\lambda) = \int \prod_{i=1}^n y_i^{\lambda_i} d\mathcal{L}(dP_{\theta_1}/dP_{\theta_o},\cdots,dP_{\theta_n}/dP_{\theta_o} \mid P_{\theta_o})(y).$$

Note that H_E does not depend on the choice of $\theta_o \in \Theta$. Since $\varphi(y) = -\prod_{i=1}^n y_i^{\lambda_i}$ is convex on $(0,\infty)^n$, it follows immediately from the criterion (1.3) that $E \geq F$ implies $H_E \leq H_F$.

Now let Θ be an open subset of $\mathbb{R}^r$ or a (not necessarily open, nondegenerate) subinterval of $\mathbb{R}$. For "smooth" experiments which are not comparable with respect to the "global" ordering $\geq$ a local comparison may be possible. If the map $\Theta \to L^1(P_{\theta_o})$, $\theta \to f_{\theta_o,\theta}$ is differentiable at $\theta_o \in \Theta$ with $f_{\theta_o,\theta} = dP_\theta/dP_{\theta_o}$, i.e. there is $\dot f_{\theta_o} \in L^1(P_{\theta_o})^r$ such that

$$\int |f_{\theta_o,\theta_o+\delta} - 1 - \delta^t \dot f_{\theta_o} \mid dP_{\theta_o} = o((\delta^t\delta)^{1/2}) \text{ as } \delta \to 0 \ \ (\delta \in \mathbb{R}^r),$$

where t denotes transposition, then E is said to be $L^1(P_{\theta_o})$-differentiable with derivative $\dot f_{\theta_o}$. Clearly, $\int \dot f_{\theta_o,i} dP_{\theta_o} = 0$ for every $i = 1,\cdots,r$. For $f \in L^1(P_{\theta_o})$, let fP_{θ_o} denote the signed measure with P_{θ_o}-density f.

Definition 2. Let E be $L^1(P_{\theta_o})$-differentiable with derivative $\dot f_{\theta_o}$ and let F be $L^1(Q_{\theta_o})$-differentiable with derivative $\dot g_{\theta_o}$, $\theta_o \in \Theta$. Then E is said to be locally at least as informative as F at θ_o, $E \geq_{\theta_o} F$, if there is a Markov kernel

M from $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$ to $(\mathcal{Y}, \mathcal{B}(\mathcal{Y}))$ such that

$$MP_{\theta_o} = Q_{\theta_o} \text{ and } M(\dot{f}_{\theta_o,i}P_{\theta_o}) = \dot{g}_{\theta_o,i}Q_{\theta_o} \text{ for every } i = 1, \cdots, r. \tag{1.5}$$

In other words, $E \geq_{\theta_o} F$ if and only if the finite "generalized" experiment $(\mathcal{X}, \mathcal{B}(\mathcal{X}), (P_{\theta_o}, \dot{f}_{\theta_o,i}P_{\theta_o} : i = 1, \cdots, r))$ is at least as informative as $(\mathcal{Y}, \mathcal{B}(\mathcal{Y}), (Q_{\theta_o}, \dot{g}_{\theta_o,i}Q_{\theta_o} : i = 1, \cdots, r))$. Analogous to the criterion (1.3) for the ordering $\geq$, the inequality $E \geq_{\theta_o} F$ is equivalent to

$$\begin{gathered} \int \varphi d\mathcal{L}(\dot{f}_{\theta_o} \mid P_{\theta_o}) \geq \int \varphi d\mathcal{L}(\dot{g}_{\theta_o} \mid Q_{\theta_o}) \\ \text{for every convex function } \varphi : \mathbb{R}^r \to \mathbb{R} \end{gathered} \tag{1.6}$$

(c.f. [28], Corollary 4.5). Since $MP_\theta = Q_\theta$ for every $\theta \in \Theta$ and some Markov kernel M implies $M(\dot{f}_{\theta_o,i}P_{\theta_o}) = \dot{g}_{\theta_o,i}Q_{\theta_o}$ for every i, $E \geq F$ implies $E \geq_{\theta_o} F$ and by the criterion (1.6), the local ordering is stronger than the usual ordering of the Fisher information matrices.

Finally, we will be concerned with the following extension of Definition 1. For probability measures Q_1 and Q_2 on $\mathcal{B}(\mathcal{Y})$, let $\|Q_1 - Q_2\| = 2\sup_{C \in \mathcal{B}(\mathcal{Y})} |Q_1(C) - Q_2(C)|$ be the variational distance.

Definition 3. The deficiency $\delta(E, F)$ of E relative to F is the number

$$\delta(E, F) = \inf_M \sup_{\theta \in \Theta} \|MP_\theta - Q_\theta\|, \tag{1.7}$$

where M ranges over all Markov kernels M from $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$ to $(\mathcal{Y}, \mathcal{B}(\mathcal{Y}))$.

Note that $0 \leq \delta(E, F) \leq 2$ and $\delta(E, F) = 0$ if and only if $E \geq F$. The decision theoretic interpretation of $\delta(E, F) \leq \epsilon$ is the same as for $E \geq F$ with (1.4) replaced by

$$\begin{aligned} &\int_{\mathcal{X}} \int_{\mathcal{D}} \ell(\theta, a)\sigma(x, da)dP_\theta(x) \\ &\qquad \leq \int_{\mathcal{Y}} \int_{\mathcal{D}} \ell(\theta, a)\rho(y, da)dQ_\theta(y) + \epsilon \sup_{a \in D} |\ell(\theta, a)|, \;\; \theta \in \Theta. \end{aligned}$$

The remainder of the paper is organized as follows. Section 2 is concerned with the ordering $\geq$ and the ordering by Hellinger transforms of general Gaussian models (1.1). The latter ordering coincides with the ordering by simple testing problems. Section 3 is devoted to linear models. Here the ordering $\geq$ coincides with the ordering by Hellinger transforms and may be completely described by linear estimation problems. In Section 4 we discuss the local ordering of nonlinear parametric models. This ordering coincides with the usual ordering of the Fisher information matrices. Finally, in Section 5, we consider the information contained in additional observation periods for parametric regression models. This quantity is measured by the deficiency. Several examples which illustrate the theory are discussed.

2. General Gaussian Regression Models. Consider models of the type (1.1). The proper setting for our results is the notion of a reproducing kernel Hilbert space (RKHS) since all information about the error process Z is contained in the RKHS of its covariance function K. Recall that K is a symmetric nonnegative definite kernel on $I \times I$. The RKHS with reproducing kernel K is a Hilbert space $H(K)$ of real valued functions on I with scalar product $\langle \cdot, \cdot \rangle_K$ and corresponding $\| \cdot \|_K$ such that:

For each $t \in I$, $K(\cdot, t)$ belongs to $H(K)$.
For $h \in H(K)$, $\langle h, K(\cdot, t)\rangle_K = h(t)$ for every $t \in I$.

Throughout we assume that

$$Z \text{ has continuous sample paths.}$$

Then $H(K)$ is contained in the separable Banach space $C(I)$ of all continuous real valued functions on I equipped with the supremum norm and the inclusion map is continuous. Assume further that

$$S(\theta) \in H(K) \text{ for every } \theta \in \Theta.$$

This ensures that the distribution $P_\theta = \mathcal{L}(X \mid \theta)$ of X under θ is equivalent to $P = \mathcal{L}(Z)$. Here $P_\theta = P \star \epsilon_{S(\theta)}$ and P are Gaussian measures on the Borel σ-algebra $\mathcal{B}(C(I))$ of $C(I)$. Thus the model (1.1) which is determined by S and K (and I) corresponds to the homogeneous experiment

$$E(S,K) = (C(I), \mathcal{B}(C(I)), (P_\theta : \theta \in \Theta)). \tag{2.1}$$

By the well known Cameron-Martin formula

$$dP_\theta/dP = \exp(L_K(S(\theta)) - \|S(\theta)\|_K^2/2), \tag{2.2}$$

where $L_K : H(K) \to L^2(P)$ denotes the linear isometry uniquely determined by

$$L_K\left(\int_I K(\cdot, t) d\mu(t)\right)(x) = \int_I x \, d\mu \quad P-\text{a.s.} \tag{2.3}$$

for every $\mu \in C(I)^*$. The topological dual $C(I)^*$ of $C(I)$ coincides with the space of all signed finite Borel measures on I. Under each P_θ, the process $L_K(h)$, $h \in H(K)$, is Gaussian with

$$E_\theta L_K(h) = \langle S(\theta), h\rangle_K, \quad \mathrm{Cov}_\theta(L_K(h_1), L_K(h_2)) = \langle h_1, h_2\rangle_K. \tag{2.4}$$

For more details on these subjets see Kuo [11] and Parzen [20].

The information in the experiment $E(S,K)$ is completely contained in the covariance function of the log-likelihood process $\log(dP_\theta/dP_{\theta_o})$, $\theta \in \Theta$, for some $\theta_o \in \Theta$, where by (2.2)

$$\log(dP_\theta/dP_{\theta_o}) = L_K(S(\theta) - S(\theta_o)) - \|S(\theta)\|_K^2/2 + \|S(\theta_o)\|_K^2/2.$$

The reason is that this process is a sufficient statistic for $E(S,K)$ and under each P_γ by (2.4), it is a Gaussian process with covariance function R_{θ_o} and mean value function given by

$$\begin{aligned} R_{\theta_o}(\theta_1,\theta_2) &= \langle S(\theta_1) - S(\theta_o), S(\theta_2) - S(\theta_o)\rangle_K, \\ \theta &\to R_{\theta_o}(\theta,\gamma) - R_{\theta_o}(\theta,\theta)/2, \end{aligned} \tag{2.5}$$

respectively.

Now let $E(S_1,K_1)$ and $E(S_2,K_2)$ be two models with corresponding covariances R_{1,θ_o} and R_{2,θ_o}. For arbitrary symmetric kernels or matrices R_1 and R_2, $R_1 \geq R_2$ is short for: $R_1 - R_2$ is nonnegative definite. It is easily verified that $R_{1,\theta_o} \geq R_{2,\theta_o}$ for some $\theta_o \in \Theta$ implies this relation for every $\theta_o \in \Theta$. Observe further that $\exp(R_{j,\theta_o})$ are symmetric nonnegative definite kernels on $\Theta \times \Theta$.

A tempting conjecture is that the ordering $\geq$ of regression models is characterized by the ordering $\geq$ of the corresponding covariances. This will be verified for several special models and, in particular, for linear models. However, for the general case we can prove only a weaker version. Here is the basic result.

THEOREM 1.

(a) *If $R_{1,\theta_o} \geq R_{2,\theta_o}$ for some $\theta_o \in \Theta$, then $E(S_1,K_1) \geq E(S_2,K_2)$.*

(b) *If $E(S_1,K_1) \geq E(S_2,K_2)$, then $\exp(R_{1,\theta_o}) \geq \exp(R_{2,\theta_o})$ for every $\theta_o \in \Theta$ and in particular, $R_{1,\theta_o}(\theta,\theta) \geq R_{2,\theta_o}(\theta,\theta)$ for every θ_o, $\theta \in \Theta$.*

(c) *$E(S_1,K_1) \sim E(S_2,K_2)$ if and only if $R_{1,\theta_o} = R_{2,\theta_o}$ for some (every) $\theta_o \in \Theta$. This is also equivalent to $R_{1,\theta_o}(\theta,\theta) = R_{2,\theta_o}(\theta,\theta)$ for every θ_o, $\theta \in \Theta$.*

PROOF.

(a) Let $\Gamma = \{\theta_1,\cdots,\theta_n\} \subset \Theta$. We have to show that $E(S_1,K_1)_\Gamma \geq E(S_2,K_2)_\Gamma$. Suppressing the index and dealing with $E(S,K)_\Gamma$, let

$$\begin{aligned} U_i &= \log(dP_{\theta_i}/dP_{\theta_o}) + R_{\theta_o}(\theta_i,\theta_i)/2 \\ &= L_K(S(\theta_i) - S(\theta_o)) - \langle S(\theta_o), S(\theta_i) - S(\theta_o)\rangle_K. \end{aligned}$$

By the Halmos-Savage criterion, $U = (U_1, \cdots, U_n) : C(I) \to \mathbb{R}^n$ is a sufficient statistic for the $E(S,K)_\Gamma$. Under P_{θ_i}, the covariance matrix of the normally distributed random vector U is given by $\Sigma = (R_{\theta_o}(\theta_j, \theta_k))_{j,k=1,\cdots,n}$ and its mean by the ith column of Σ according to (2.5). Hence, $\mathcal{L}(U \mid P_{\theta_i}) = N(\Sigma e_i, \Sigma)$, where e_i denotes the ith unit vector in $\mathbb{R}^n$. Identifying Γ with $\{1, \cdots, n\}$ this yields

$$E(S,K)_\Gamma \sim (N(\Sigma e_i, \Sigma) : i \in \Gamma).$$

So the reduction by sufficiency leads to the conclusion that $E(S_1, K_1)\Gamma \geq E(S_2, K_2)\Gamma$ if and only if

$$(N(\Sigma_1 e_i, \Sigma_1) : i \in \Gamma) \geq (N(\Sigma_2 e_i, \Sigma_2) : i \in \Gamma), \tag{2.6}$$

where Σ_j is defined with respect to R_{j,θ_o}. Now, by assumption, $\Sigma_1 \geq \Sigma_2$. Let $\Pi_j : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n$, $\Pi_j(y_1, y_2) = y_j$ and $Q_i = N((\Sigma_1 - \Sigma_2)e_i, \Sigma_1 - \Sigma_2) \otimes N(\Sigma_2 e_i, \Sigma_2)$. Then $\Pi_1 + \Pi_2$ is sufficient for the experiment $(Q_i : i \in \Gamma)$ and defining the Markov kernel M on $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ by $M(x) = Q_i^{\Pi_2 | \Pi_1 + \Pi_2 = x}$ independent of $i \in \Gamma$, we obtain $MN(\Sigma_1 e_i, \Sigma_1) = N(\Sigma_2 e_i, \Sigma_2)$ for every $i \in \Gamma$ which yields (2.6). The implication $\Sigma_1 \geq \Sigma_2 \Rightarrow$ (2.6) is a special case of the characterization of $\geq$ for classical linear models, see Example 3 for references.

We give a second proof based on the convex function criterion (1.3).

SECOND PROOF. Let $\Gamma = \{\theta_o, \theta_1, \cdots, \theta_n\} \subset \Theta$ and let $\widetilde{P}_j$ denote the distribution on $(0,\infty)^n$ of the likelihood process $dP_{j,\theta_i}/dP_{j,\theta_o}$, $i = 1, \cdots, n$, under P_{j,θ_o}, $j = 1, 2$. By (2.2) and (2.4), $\widetilde{P}_j = \mathcal{L}(\exp \mid N(a_j, \Sigma_j))$, where $\exp : \mathbb{R}^n \to (0,\infty)^n$, $\exp y = (\exp y_1, \cdots, \exp y_n)$, Σ_j is defined as above and $a_j = -(R_{j,\theta_o}(\theta_1, \theta_1), \cdots, R_{j,\theta_o}(\theta_n, \theta_n))^t/2$. Therefore, by the convex function criterion (1.3), $E(S_1, K_1)_\Gamma \geq E(S_2, K_2)_\Gamma$ if and only if

$$\int \varphi d\mathcal{L}(\exp \mid N(a_1, \Sigma_2)) \geq \int \varphi d\mathcal{L}(\exp \mid N(a_2, \Sigma_2)) \quad \text{for every convex function } \varphi : (0,\infty)^n \to \mathbb{R}. \tag{2.7}$$

Since $\Sigma_1 \geq \Sigma_2$, we can define a Markov kernel M on $((0,\infty)^n, \mathcal{B}((0,\infty)^n))$ by

$$M(y) = \mathcal{L}(\exp \mid N(\log y + a_1 - a_2, \Sigma_1 - \Sigma_2)),$$

where $\log : (0,\infty)^n \to \mathbb{R}^n$, $\log y = (\log y_1, \cdots, \log y_n)^t$. Using the formula for the Laplace transform of a normal distribution, we see that

$$\begin{aligned} \int x_i M(y, dx) &= \exp\{e_i^t(\log y + a_1 - a_2) + e_i^t(\Sigma_1 - \Sigma_2)e_i/2\} \\ &= y_i \text{ for every } y \in (0,\infty)^n, \end{aligned}$$

that is, M is a dilation. Moreover, for $A \in \mathcal{B}((0,\infty)^n)$ we have

$$\begin{aligned} M\widetilde{P}_2(A) &= \int M(\exp x, A)dN(a_2, \Sigma_2)(x) \\ &= \int N(x + a_1 - a_2, \Sigma_1 - \Sigma_2)(\exp^{-1}(A))dN(a_2, \Sigma_2)(x) = \widetilde{P}_1(A). \end{aligned}$$

Now (2.7) follows from Jensen's inequality.

(b) Let $\theta_o \in \Theta$, $\Gamma = \{\theta_1, \cdots, \theta_n\} \subset \Theta$ and $\alpha_1, \cdots, \alpha_n \in \mathbb{R}$. It suffices to consider the case $\theta_o \notin \Gamma$. By assumption and the convex function criterion, (2.7) holds. For φ defined by $\varphi(y) = (\sum_{i=1}^n \alpha_i y_i)^2$ and $j = 1, 2$, one gets

$$\begin{aligned} \int \varphi d\widetilde{P}_j &= \sum_{i,k=1}^n \alpha_i \alpha_k \int \exp(x_1 + x_k) dN(a_j, \Sigma_j)(x) \\ &= \sum_{i,k=1}^n \alpha_i \alpha_k \exp\{(e_i + e_k)^t a_j + (e_i + e_k)^t \Sigma_j (e_i + e_k)/2\} \\ &= \sum_{i,k=1}^n \alpha_i \alpha_k \exp(R_{j,\theta_o}(\theta_i, \theta_k)). \end{aligned}$$

This proves the assertion.

(c) This part is known (cf. e.g. [18], Chapter I.2). The implications $R_{1,\theta_o} = R_{2,\theta_o}$ for some $\theta_o \in \Theta \Rightarrow E(S_1, K_1) \sim E(S_2, K_2) \Rightarrow R_{1,\theta_o}(\theta, \theta) = R_{2,\theta_o}(\theta, \theta)$ for every $\theta_o, \theta \in \Theta$ are immediate from (a) and (b). Moreover, for $\theta_o \in \Theta$, $j = 1, 2$

$$R_{j,\theta_o}(\theta, \gamma) = (R_{j,\theta_o}(\theta, \theta) + R_{j,\theta_o}(\gamma, \gamma) - R_{j,\gamma}(\theta, \theta))/2$$

for every $\theta, \gamma \in \Theta$ which yields the assertions. ∎

In view of

$$\exp(R_{j,\theta_o}(\theta_1, \theta_2)) = \int (dP_{j,\theta_1}/dP_{j,\theta_o})(dP_{j,\theta_2}/dP_{j,\theta_o})dP_{j,\theta_o}, \quad \theta_1, \theta_2 \in \Theta, \quad (2.8)$$

one can show that the necessary condition given in part (b) coincides with the ordering by the variances of locally MVU estimators, namely, $\exp(R_{1,\theta_o}) \geq \exp(R_{2,\theta_o})$ if and only if each functional $f : \Theta \to \mathbb{R}$ which admits an unbiased $L^2(P_{2,\theta_o})$-estimator in $E(S_2, K_2)$ also admits an unbiased $L^2(P_{1,\theta_o})$-estimator in $E(S_1, K_1)$ and $\mathrm{Var}_{\theta_o} g_1 \leq \mathrm{Var}_{\theta_o} g_2$, where g_j is the locally MVUE of f at θ_o based on $E(S_j, K_j)$.

EXAMPLE 1. Let $X(t) = Ah(\theta t) + Z(t)$, $t \in I$ (e.g. $h(t) = \sin t$) with $\Theta \subset (0,\infty)$ and $A > 0$ such that $t \to h(\theta t) \in H(K)$ for every $\theta \in \Theta$. Setting $S_A(\theta, t) = Ah(\theta t)$, $E(S_{A_1}, K) \geq E(S_{A_2}, K)$ holds if and only if $A_1 \geq A_2$ provided that $|\{t \to h(\theta t) : \theta \in \Theta\}| \geq 2$. So not surprisingly, the information about the frequency in a (e.g. sinusoidal) signal corrupted by additive Gaussian noise is an increasing function of the amplitude.

EXAMPLE 2. (Gaussian diffusion processes). Consider Gaussian diffusion models

$$dX(t) = (A(t)X(t) + \theta(t)dt + \epsilon dW(t), \;\; X(0) = 0, \;\; t \in [0,T], \;\; T < \infty, \quad (2.9)$$

where $A \in L^1([0,T], dt)$, $\epsilon > 0$, $\theta \in \Theta \subset L^2([0,T], dt)$ and W denotes a standard Wiener process (with respect to a complete, right continuous filtration). Setting $P_{A,\epsilon,\theta} = \mathcal{L}(X \mid A, \epsilon, \theta)$, (2.9) yields the experiment

$$E_{A,\epsilon} = (C_o[0,T], \mathcal{B}(C_o[0,T]), \;\; (P_{A,\epsilon,\theta} : \theta \in \Theta)),$$

where $C_o[0,T] = \{x \in C[0,T] : x(0) = 0\}$. (The replacement of $C[0,T]$ by $C_o[0,T]$ does not effect the results of this section.) Then for $A_j \in L^1([0,T], dt)$ and $\epsilon_j > 0$, $j = 1,2$, we claim that $E_{A_1,\epsilon_1} \geq E_{A_2,\epsilon_2}$ if and only if $\epsilon_1 \leq \epsilon_2$ assuming $|\Theta| \geq 2$. The remarkable fact is that the ordering $\geq$ does not depend on functions A_j. In particular, $E_{A_1,\epsilon}$ and $E_{A_2,\epsilon}$ are equally informative.

The unique (strong) solution of (2.9) is

$$X(t) = \psi_A(t) \left\{ \int_o^t \psi_A(s)^{-1}\theta(s)ds + \epsilon \int_o^t \psi_A(s)^{-1} dW(s) \right\},$$

where

$$\psi_A(t) = \exp\left(\int_o^t A(s)ds \right)$$

(cf. [14], Theorem 4.10). The process $Z(t) = \epsilon\psi_A(t)\int_o^t \psi_A(s)^{-1}dW(s)$ is a zero-mean Gaussian process with covariance function

$$K_{A,\epsilon}(s,t) = \epsilon^2 \psi_A(s)\psi_A(t) \int_o^{\min(s,t)} \psi_A(r)^{-2} dr. \quad (2.10)$$

Defining $S_A(\theta, t) = \psi_A(t)\int_o^t \psi_A(s)^{-1}\theta(s)ds$, we see that $E_{A,\epsilon} = E(S_A, K_{A,\epsilon})$. Moreover, the RKHS of $K_{A,\epsilon}$ is given by

$$H(K_{A,\epsilon}) = \{h \in C_o[0,T] : h(t) = \epsilon\psi_A(t) \int_o^t \psi_A(s)^{-1} g(s)ds, \;\; g \in L^2([0,T], dt)\}$$

equipped with the scalar product

$$\langle h_1, h_2\rangle_{K_{A,\epsilon}} = \int_o^T g_1(t)g_2(t)dt. \tag{2.11}$$

This follows from [20], Theorem 4D, p. 297. Therefore, $S_{A,\epsilon}(\theta) \in H(K_{A,\epsilon})$ for every $\theta \in \Theta$ and

$$R_{A,\epsilon,\theta_o}(\theta_1, \theta_2) = \epsilon^{-2} \int_o^T (\theta_1(t) - \theta_o(t))(\theta_2(t) - \theta_o(t))dt.$$

Our claim now follows from Theorem 1.

For instance, in case $A = 0$, $S_0(\theta, t) = \int_0^t \theta(s)ds$ and $K_{o,\epsilon}(s,t) = \epsilon^2 \min(s,t)$ and for $A(t) = (t-c)^{-1}$, $T < c$, we have $S_A(\theta, t) = \frac{c-t}{c}\int_o^t \frac{c}{c-s}\theta(s)ds$ and $K_{A,\epsilon}(s,t) = \epsilon^2(\min(s,t) - st/c)$ (Brownian bridge on $[o,c]$). The estimation problem for functionals of θ in the model $E_{o,\epsilon} = E(S_o, K_{o,\epsilon})$ for various sets Θ has been treated by Ibragimov et al. [6], Section VII.4, [7], [8].

The following consequence of Theorem 1 is of interest for regression models with a large set of possible regression functions.

COROLLARY 2. *Suppose that $S_1 = S_2 = S$ ($I_1 = I_2 = I$) and $S(\Theta)$ is a dense subset of $H(K_1)$. Then the following conditions are equivalent:*

(I) $E(S, K_1) \geq E(S, K_2)$.

(II) $K_1 \leq K_2$.

(III) $H(K_1) \subset H(K_2)$ *and* $\|h\|_{K_1} \geq \|h\|_{K_2}$ *for every* $h \in H(K_1)$.

(IV) $R_{1,\theta_o} \geq R_{2,\theta_o}$ *for some* (every) $\theta_o \in \Theta$.

(V) $R_{1,\theta_o}(\theta, \theta) \geq R_{2,\theta_o}(\theta, \theta)$ *for every* θ_o, $\theta \in \Theta$.

PROOF. (II) $\Leftrightarrow$ (III) is well known (cf. [10], Theorems 3.3, 3.4) and (IV) $\Rightarrow$ (I) $\Rightarrow$ (V) is contained in Theorem 1. (III) $\Rightarrow$ (IV) follows immediately from $\sum_{i,k} \alpha_i\alpha_k R_{j,\theta_o}(\theta_i, \theta_k) = \| \sum_i \alpha_i(S(\theta_i) - S(\theta_o))\|^2_{K_j}$, $\alpha_i \in \mathbb{R}$.

(V) $\Rightarrow$ (III). We prove first that $H(K_1) \subset H(K_2)$. Let $h \in H(K_1)$. Choose a sequence θ_n, $n \in \mathbb{N}$, in Θ such that $\|S(\theta_n) - h\|_{K_1} \to 0$. Since by (V), $\|S(\theta_n) - S(\theta_m)\|_{K_2} \leq \|S((\theta_n) - S(\theta_m)\|_{K_1}$, $S(\theta_n)$, $n \in \mathbb{N}$, is a Cauchy sequence in $H(K_2)$ and hence, $\|S(\theta_n) - g\|_{K_2} \to 0$ for some $g \in H(K_2)$. Since the inclusion maps $H(K_j) \to C(I)$ are continuous, $S(\theta_n) \to h$ and $S(\theta_n) \to g$ in the norm topology of $C(I)$. Thus $h = g$ which yields $h \in H(K_2)$. Next observe that the inclusion map $H(K_1) \to H(K_2)$ is continuous by the closed graph theorem and $S(\Theta) - S(\Theta)$ is a dense subset of $H(K_1)$. Therefore, (V) implies $\|h\|_{K_1} \geq \|h\|_{K_2}$ for ever $h \in H(K_1)$ and the proof of (III) is complete.∎

Thus for the specific models of the preceding corollary, the ordering $\geq$ coincides with the ordering $\leq$ of the covariance functions of the error processes and also with the ordering by Hellinger transforms as the following characterization shows.

PROPOSITION 3. *The following statements are equivalent:*

(I) The Hellinger transform of $E(S_1, K_1)$ is pointwise $\leq$ the Hellinger transform of $E(S_2, K_2)$.

(II) $E(S_1, K_1)_\Gamma \geq E(S_2, K_2)_\Gamma$ for every subset Γ of Θ with $|\Gamma| = 2$.

(III) $R_{1,\theta_o}(\theta, \theta) \geq R_{2,\theta_o}(\theta, \theta)$ for every θ_o, $\theta \in \Theta$.

PROOF. (I) $\Leftrightarrow$ (III). The Hellinger transform of $E(S, K)$ is given by

$$\exp\left(-\sum_{\theta,\gamma} \lambda(\theta)\lambda(\gamma)\|S(\theta) - S(\gamma)\|_K^2/4\right)$$

for every prior distribution λ on Θ with finite support.

(II) $\Leftrightarrow$ (III). In case $|\Theta| = 2$, the condition $R_{1,\theta_o} \geq R_{2\theta_o}$ is equivalent to $R_{1,\theta_o}(\theta, \theta) \geq R_{2,\theta_o}(\theta, \theta)$ for every $\theta \in \Theta$ since $R_{j,\theta_o}(\theta_o, \theta) = 0$. Thus the assertion follows immediately from Theorem 1. ∎

An immediate implication is the following interpretation by simple testing problems of the ordering by Hellinger transforms. For $\Gamma = \{\theta_1, \theta_2\} \subset \Theta$ and $\alpha \in (0, 1)$ let

$$\beta_\Gamma(\alpha) = \sup\left\{\int \varphi dP_{\theta_2} : \int \varphi dP_{\theta_1} \leq \alpha, \ \ \varphi \ \text{ test}\right\}$$

be the power of the most powerful level α test for testing $\theta = \theta_1$ vs. $\theta = \theta_2$ in $E(S, K)$. By the Neyman-Pearson lemma,

$$\beta_\Gamma(\alpha) = \Phi(-z_\alpha + \|S(\theta_1) - S(\theta_2)\|_K),$$

where Φ denotes the df of $N(0, 1)$ and $\Phi(z_\alpha) = 1 - \alpha$, so that condition (III) is equivalent to $\beta_{1,\Gamma}(\alpha) \geq \beta_{2,\Gamma}(\alpha)$ for every $\alpha \in (0, 1)$ and $\Gamma \subset \Theta$ with $|\Gamma| = 2$. The equivalence of the latter condition and (II) follows also from [23], 15.0 and 54.2.

REMARKS.

(a) Without any change the results are valid for compact subsets I of $\mathbb{R}^k$, $k > 1$. Therefore, it would be possible to deal with Gaussian random fields.

(b) The topological support $\operatorname{supp}(P)$ of $P = \mathcal{L}(Z)$ is a closed linear subspace of $C(I)$. One may choose as sample space for $E(S, K)$ any closed linear subspace of $C(I)$ containing $\operatorname{supp}(P)$ instead of $C(I)$ as we did in (2.1).

3. Linear Models. In this section we deal with models (1.1) when Θ is a linear space and S is a linear map. In this case, the covariance of the log-likelihood process (w.r.t. $\theta_o = 0$) is bilinear. Hence, the converse of Theorem 1.a is valid and the ordering $\geq$ may be characterized as follows. Recall that $||S(\theta)||_K^2$ is the variance of $\log(dP_\theta/dP_o)$.

PROPOSITION 4. *Let Θ be a linear space and let $S_j : \Theta \to H(K_j)$ be linear maps, $j = 1, 2$. Then $E(S_1, K_1) \geq E(S_2, K_2)$ if and only if*

$$||S_1(\theta)||_{K_1} \geq ||S_2(\theta)||_{K_2} \text{ for every } \theta \in \Theta. \tag{3.1}$$

PROOF. In view of the assumptions, condition (3.1) is equivalent to $R_{1,0} \geq R_{2,0}$. Since (3.1) is also equivalent to the condition $R_{1,0}(\theta, \theta) \geq R_{2,0}(\theta, \theta)$ for every $\theta \in \Theta$, the assertion follows from Theorem 1. ∎

This proposition combined with Proposition 3 shows that the ordering $\geq$ of linear models coincides with the ordering by Hellinger transforms and by binary subexperiments. Moreover, it coincides with the ordering by variance of BLUE's (or UMVUE's) for linear functionals of θ. Here estimators $L_k(h)$, $h \in H(K)$, are called linear estimators in $E(S, K)$. It is not difficult to check that the space of these estimators is the $L^2(P)$-closure of the space of usual linear estimators $x \to \int_I x \, d\mu$, $\mu \in C(I)^*$, see (2.3). By (2.4), the variance of linear estimators does not depend on θ.

PROPOSITION 5. *Let Θ be a linear space and let $S_j : \Theta \to H(K_j)$ be linear, $j = 1, 2$. Then $E(S_1, K_1) \geq E(S_2, K_2)$ if and only if each linear functional $f : \Theta \to \mathbb{R}$ which admits a linear unbiased estimator in $E(S_2, K_2)$ also admits a linear unbiased estimator in $E(S_1, K_1)$ and $\operatorname{Var} g_1 \leq \operatorname{Var} g_2$, where g_j is the BLUE for $f(\theta)$ based on $E(S_j, K_j)$. Furthermore, the BLUE is even UMVU.*

PROOF. By Proposition 4, $E(S_1, K_1) \geq E(S_2, K_2)$ if and only if $R_{1,0} \geq R_{2,0}$. The latter condition is equivalent to $H(R_{1,0}) \supset H(R_{2,0})$ and $||f||_1 \leq ||f||_2$ for every $f \in H(R_{2,0})$, where $||\cdot||_j$ denotes the norm in $H(R_{j,0})$ (cf. [10], Theorems 3.3, 3.4). In view of (2.5), we have

$$H(R_{j,0}) = \{f : \Theta \to \mathbb{R} : f(\theta) = \langle h, S_j(\theta) \rangle_{K_j}, \ h \in \overline{S_j(\Theta)}^{(K_j)}\}$$

equipped with the scalar product

$$\langle f_1, f_2 \rangle_j = \langle h_1, h_2 \rangle_{K_j},$$

where $\overline{S_j(\Theta)}^{(K_j)}$ denotes the closure of $S_j(\Theta)$ in $H(K_j)$ (cf. [20], Theorem 7B, p. 325). Furthermore, the BLUE for $f(\theta) = \langle h, S_j(\theta)\rangle_{K_j}$, $h \in H(K_j)$, is given by $g_j = L_{K_j}(h^*)$, where h^* is the orthogonal projection of h onto $\overline{S_j(\Theta)}^{(K_j)}$, and Var $g_j = \|h^*\|^2_{K_j} = \|f\|^2_j$. This proves the desired characterization of $\geq$.

Suppressing the index j, let $L_K(h^*)$ be the BLUE for $f(\theta) = \langle h^*, S(\theta)\rangle_K$, $h^* \in \overline{S(\Theta)}^{(K)}$. Choose an orthonormal basis, h_n, $n \geq 1$, of $\overline{S(\Theta)}^{(K)}$ contained in $S(\Theta)$ (note that $H(K)$ is separable) and let $\mathcal{A}$ be the sub-σ-algebra of $\mathcal{B}(C(I))$ generated by the sequence $L_K(h_n)$, $n \geq 1$, and the P_o-null sets. Then $L_K(h)$ is $\mathcal{A}$-measurable for every $h \in \overline{S(\Theta)}^{(K)}$. Hence, by (2.2) and the Halmos-Savage criterion, $\mathcal{A}$ is a sufficient σ-algebra for $E(S, K)$. Moreover, $(P_\theta \mid \mathcal{A} : \theta \in \Theta)$ is complete. To see this, let $u \in \cap_{\theta\in\Theta} L^1(P_\theta \mid \mathcal{A})$ such that $E_\theta u = 0$ or every $\theta \in \Theta$, that is, $E_o[u \exp(L_K(h) - \|h\|^2_K/2)] = 0$ for every $h \in S(\Theta)$. For every h in the linear span of $\{h_1, \cdots, h_n\}$, we obtain

$$E_o\left[E_o(u \mid L_K(h_1), \cdots, L_k(h_n)) \exp\left(\sum_{i=1}^{n} \langle h, h_i\rangle_K L_K(h_i) - \sum_{i=1}^{n} \langle h, h_i\rangle^2_K/2\right)\right] = 0,$$

which yields $E_o(u \mid L_K(h_1), \cdots, L_K(h_n)) = 0$ P_o-a.s. and from the martingale convergence theorem follows $E_o(u \mid \mathcal{A}) = u = 0$ P_o-a.s. Now we can conclude that $L_k(h^*)$, being $\mathcal{A}$-measurable, is UMVU for $f(\theta)$ by the theorem of Lehmann-Scheffé. ∎

EXAMPLE 3. (Classical linear models). We can deduce from Proposition 4 the following known result on the comparison of classical linear models, see Stepniak and Torgersen [22] and Torgersen [27] (cf. also [3], [5], [9]). Let $E_j = (\mathbb{R}^{n_j}, \mathcal{B}(\mathbb{R}^{n_j}), (N(A_j\theta, \Sigma_j) : \theta \in \mathbb{R}^r))$, where A_j is an $n_j \times r$ matrix, Σ_j is a (possibly singular) nonnegative definite $n_j \times n_j$ matrix and assume that $\{A_j\theta : \theta \in \mathbb{R}^r\} \subset \{\Sigma_j y : y \in \mathbb{R}^{n_j}\}$, $j = 1, 2$. If we set $I_j = \{1, \cdots, n_j\}$, $K_j(i, k) = e_i\Sigma_j e_k$, $S_j(\theta) = A_j\theta$ and identify $C(I_j)$ with $\mathbb{R}^{n_j}$, then $H(K_j) = \{\Sigma_j y : y \in \mathbb{R}^{n_j}\}$ equipped with the scalar product

$$\langle x_1, x_2\rangle_{K_j} = x_1^t \overline{\Sigma}_j x_2, \tag{3.2}$$

where $\overline{\Sigma}_j$ denots an arbitrary generalized inverse of Σ_j, and $E_j = E(S_j, K_j)$ with $\Theta = \mathbb{R}^r$. Note that (3.2) does not depend on the choice of $\overline{\Sigma}_j$. We thus obtain $E_1 \geq E_2$ if and only if $A_1^t \overline{\Sigma}_1 A_1 \geq A_2^t \overline{\Sigma} A_2$.

For $a, b \in \mathbb{R}$, $a < b$, let

$$W_1^2[a, b] = \{h : [a, b] \to \mathbb{R} : h \text{ is absolutely continuous } h' \in L^2([a, b], dt)\}.$$

In all subsequent examples, the RKHS's are Sobolev spaces of the above type equipped with a suitable scalar product.

EXAMPLE 4. (Gaussian white noise and colored noise). Consider the model

$$X(t) = \theta(t) + N(t), \qquad t \in [0,T],$$

where N denotes Gaussian white noise (with intensity 1) and $\Theta = W_1^2[0,T]$. If the ordering $\geq$ is a reasonable concept, then this model should be less informative than the model where N is replaced by the colored (stationary) noise with covariance function $K_1(s,t) = \exp(-|s-t|)/2$. In fact, this is true. Following the conventional approach to deal rigorously with Gaussian white noise we replace the above equation by the integrated version

$$Y(t) = \int_o^t \theta(s)ds + W(t), \qquad t \in [0,T],$$

where W is a standard Wiener process. Setting $S_1(\theta) = \theta$, $S_2(\theta,t) = \int_o^t \theta(s)ds$ and $K_2(s,t) = \min(s,t)$, we claim that $E(S_1,K_1) \geq E(S_2,K_2)$ ($I_1 = I_2 = [0,T]$). Note first that $H(K_1) = W_1^2[0,T]$ equipped with the scalar product

$$\langle h_1, h_2\rangle_{K_1} = 2h_1(0)h_2(0) + \int_o^T (h_1 + h_1')(h_2 + h_2')dt$$

(cf. [20], p. 430) and $H(K_2) = \{h \in W_1^2[0,T] : h(0) = 0\}$ equipped with the scalar product

$$\langle h_1, h_2\rangle_{K_2} = \int_o^T h_1'(t)h_2'(t)dt. \tag{3.3}$$

Using integration by parts this yields for $\theta \in \Theta$

$$||\theta||_{K_1}^2 = \theta^2(0) + \theta^2(T) + \int_o^T (\theta^2(t) + \theta'^2(t))dt \geq \int_o^T \theta^2(t)dt = ||S_2(\theta)||_{K_2}^2.$$

The desired informational inequality follows from Proposition 4.

In the sequel we consider some special models.

Parametric regression models. Let

$$X(t) = \sum_{i=1}^r \theta_i h_i(t) + Z(t), \qquad t \in I, \tag{3.4}$$

where $h = (h_1, \cdots, h_r) \in H(K)^r$ and $\Theta = \mathbb{R}^r$. In this model $S(\theta) = \sum_{i=1}^r \theta_i h_i$ and the corresponding experiment is denoted more concisely by $E(h,K)$. Define a symmetric nonnegative definite $r \times r$ matrix by $J(h,K) = (\langle h_i, h_j\rangle_K)_{i,j=1,\cdots,r}$. Then according to Proposition 4, given $h \in H(K_1)^r$ and $g \in H(K_2)^r$, $E(h,K_1) \geq E(g,K_2)$ if and only if

$$J(h,K_1) \geq J(g,K_2). \tag{3.5}$$

Since $J(h,K)$ is the Fisher information matrix of $E(h,K)$, the ordering $\geq$ of parametric regression models coincides with the ordering $\geq$ of the Fisher information matrices. Note further that the ML-estimator of θ in $E(h,K)$ is given by $\hat{\theta} = J(h,K)^+(L_K(h_1),\cdots,L_K(h_r))^t$ and its covariance matrix by $J(h,K)^+$ under each P_θ, where $J(h,K)^+$ denotes the Moore-Penrose inverse of $J(h,K)$. If $J(h,K)$ is regular, $\hat{\theta}$ is an unbiased estimator of θ which is UMVU in the sense that Cov $\hat{\theta} \leq \mathrm{Cov}_\theta\, \tilde{\theta}$ for every $\theta \in \Theta$ and every other unbiased estimator $\tilde{\theta} \in \cap_{\theta\in\Theta} L^2(P_\theta)$ of θ, see Proposition 5 and [29], Satz 2.114. Hence, the ordering $\geq$ of parametric regression models with independent regression functions may also be characterized by the ordering $\leq$ of the covariance matrices of the ML-estimators (UMVUE's) of θ.

Let us briefly discuss an application to the comparison of sampling designs in (3.4). Designing problems for parametric regression models with not necessarily Gaussian error processes are treated in the book of Näther [19] from a different point of view; see also the survey article of Bandemer et al. [1].

Suppose that observations $X(t)$ are available only at $t \in D$ for some closed subset D of I. Let $E((h,K)_D$, $J(h,K)_D$ and $\|g\|_D$ for $g \in H(K)$ be short for $E(h_D,K_D)$, $J(h_D,K_D)$ and $\|g_D\|_{K_D}$, respectively, where h_D denotes the restriction of h on D and K_D denotes the restriction of K on $D\times D$. Note that $g_D \in H(K_D)$ and $\|g\|_D \leq \|g\|_I$ for every $g \in H(K)$, $D \subset I$ (cf. [20], Theorem 6C, p. 312). For (closed) designs $D_1, D_2 \subset I$, it is natural to say that D_1 is at least as informative as D_2 in $E(h,K)$ if $E(h,K)_{D_1} \geq E(h,K)_{D_2}$. By (3.5), this is equivalent to the condition

$$J(h,K)_{D_1} \geq J(h,K)_{D_2}. \tag{3.6}$$

Clearly, I is at least as informative as any other design. The most interesting situations are discrete sampling and continuous sampling in closed subintervals of I. In case of discrete (finite) designs $D_1, D_2 \subset I$, (3.6) takes the form

$$h^t_{D_1}\bar{K}_{D_1}h_{D_1} \geq h^t_{D_2}\bar{K}_{D_2}h_{D_2}, \tag{3.7}$$

where for $D = \{t_1,\cdots,t_n\}$, h_D is identified with the $n\times r$ matrix $(h_j(t_i))_{\substack{i=1,\cdots,n\\ j=1,\cdots,r}}$ and K_D with the $n\times n$ matrix $(K(t_i,t_j))_{i,j=1,\cdots,n}$, see (3.2). We remark that for the comparison of discrete designs D with regular matrix K_D, the assumption $h \in H(K)^r$ can be dropped, since then $H(K_D)$ coincides with $\mathbb{R}^D$ as set. In the special case

$$h_j = \sum_{i=1}^{n} \alpha_{ji}K(\cdot,t_i) \text{ for every } j = 1,\cdots,r, \ \ \alpha_{ji} \in \mathbb{R},$$

we obviously have $J(h,K)_I = J(h,K)_D$ with $D = \{t_1, \cdots, t_n\}$, so that D and I are equally informative. This is best understood by the observation that $C(I) \to \mathbb{R}^n$, $x \to (x(t_1), \cdots, x(t_n))$ is a sufficient statistic for $E(h,K)$ in view of (2.2) and (2.3).

EXAMPLE 5. Let $r = 1$ and $K(s,t) = \min(s,t)$ on $[0,T]$. For two-point designs $D = \{t_1, t_2\}$, $0 \le t_1 < t_2 \le T$, we have

$$||h||_D^2 = h^2(t_1)/t_1 + (h(t_1) - h(t_2))^2/(t_2 - t_1)$$

if $t_1 > 0$ and $||h||_D^2 = h^2(t_2)/t_2$ otherwise. In case $h(t) = t^2$, $||h||_D^2 = t_1^3 + (t_2 - t_1)(t_2 + t_1)^2$ and thus, $D^* = \{T/2, T\}$ is at least as informative as any other two-point design. If $h(t) = t^3$, then $||h||_D^2 = t_1^5 + (t_2 - t_1)(t_1^2 + t_1t_2 + t_2^2)^2$ and $D^* = \{T(1 + 17^{1/2})/8, T\}$ is at least as informative as any other two-point design. In case $h(t) = t = K(t,T)$, $\{T\}$ and $[0,T]$ are equally informative.

Now consider continuous designs $D = [a, a+\ell] \subset [0,T]$ of fixed length ℓ, $0 < \ell < T$. For any interval $[a,b]$ with $a > 0$, we have $H(K_{[a,b]}) = W_1^2[a,b]$ equipped with the scalar product

$$\langle h_1, h_2 \rangle_{[a,b]} = a^{-1}h_1(a)h_2(a) + \int_a^b h_1'(t)h_2'(t)dt;$$

for the case $a = 0$ see (3.3). Again for $h(t) = t^2$, we obtain $||h||_D^2 = 4(a+\ell)^3/3 - a^3/3$ and for $h(t) = t^3$, $||h||_D^2 = 9(a+\ell)^5/5 - 4a^5/5$ holds. This implies that for both regression functions, $D^* = [T-\ell, T]$ is at least as informative as any other design interval of length ℓ.

Nonparametric regression models. Let

$$X(t) = \theta(t) + Z(t), \qquad t \in I, \tag{3.8}$$

where Θ is a (possibly infinite dimensional) linear subspace of $H(K)$. Here $S(\theta) = \theta$ and the corresponding experiment is simply denoted by $E(K)$. Given two such experiments $E(K_1)$, $E(K_2)$ on Θ ($I_1 = I_2$, $\Theta \subset H(K_1) \cap H(K_2)$), (3.1) takes the form

$$||\theta||_{K_1} \ge ||\theta||_{K_2} \text{ for every } \theta \in \Theta. \tag{3.9}$$

As in the proof of Corollary 2, one can show that (3.9) implies $\overline{\Theta}^{(K_1)} \subset \overline{\Theta}^{(K_2)}$ and $||\theta||_{K_1} \ge ||\theta||_{K_2}$ for every $\theta \in \overline{\Theta}^{(K_1)}$, where $\overline{\Theta}^{(K_j)}$ denotes the closure of Θ in $H(K_j)$.

Let us consider in more detail error processes which are solutions of linear stochastic differential equations of the following type

$$dZ(t) = A(t)Z(t)dt + dW(t), \ \ Z(0) = 0, \ \ t \in [0,T], \tag{3.10}$$

where A is a continuously differentiable function on $[0,T]$. The covariance function K_A of Z is given by (2.10) (with $\epsilon = 1$) and the corresponding experiment $E(K_A)$ is equipped with the sample space $C_o(0,T]$ of all continuous functions vanishing at zero. Since $\mathcal{L}(Z \mid A)$ is equivalent to $\mathcal{L}(Z \mid 0) = \mathcal{L}(W)$ (cf. [14], Theorem 7.5), the RKHS $H(K_A)$ coincides with $H(K_o)$ as set and their norms are equivalent (cf. [10], Theorems 3.3, 3.4 and 4.5). In particular, $H(K_A)$ and $H(K_o)$ coincide as topological spaces.

THEOREM 6. *Let A and B be continuously differentiable functions on $[0,T]$ and let Θ be a dense (not necessarily linear) subset of $H(K_A)$. Then $E(K_A) \geq E(K_B)$ if and only if*

$$A' + A^2 \geq B' + B^2 \ \text{ on } [0,T] \text{ and } A(T) \leq B(T). \tag{3.11}$$

PROOF. According to Corollary 2, $E(K_A) \geq E(K_B)$ holds if and only if

$$\|h\|_{K_A} \geq \|h\|_{K_B} \ \text{ for every } \ h \in H(K_A). \tag{3.12}$$

For $h \in H(K_A)$, $h(t) = \psi_A(t) \int_o^t \psi_A(s)^{-1} g(s) ds$ with $g \in L^2([0,T]), dt)$, we obtain from (2.11)

$$\|h\|^2_{K_A} = \int_o^T g^2(t)dt.$$

Using integration by parts, one gets

$$\begin{aligned}\|h\|^2_{K_A} &= \int_o^T [h'(t) - A(t)h(t)]^2 dt = \int_o^T [h'^2(t) + (A'(t) + A^2(t))h^2(t)]dt \\ &\quad - A(T)h^2(T) = \int_o^T h'^2(t)dt - \int_o^T h^2(t) d\mu_A(t),\end{aligned}$$

where $\mu_A = A(T)\epsilon_T - (A' + A^2)dt \in C_o[0,T]^*$. Therefore, (3.12) is equivalent to the condition

$$\int_o^T h^2 d(\mu_B - \mu_A) \geq 0 \ \text{ for every } \ h \in H(K_A). \tag{3.13}$$

Since the functional $C_o[0,T] \to \mathbb{R}$, $x \to \int_o^T x^2 d(\mu_B - \mu_A)$ is continuous and $H(K_A)$ is dense in $C_o[0,T]$, (3.13) holds if and only if $\mu_B - \mu_A \geq 0$ and this condition is equivalent to (3.11). Thus the proof is complete. ∎

REMARKS.

(a) In case $A = 0$ or $B = 0$, the preceding theorem has been proved in [16]. If Θ is a linear subspace of $H(K_A)$, it is enough to impose the (weaker) assumption that Θ is dense in $C_o[0,T]$. To see this, one only has to replace Corollary 2 by the criterion (3.9) in the proof of Theorem 6.

(b) Using (3.11) it is easy to construct functions A and B such that $E(K_A) \geq E(K_B)$ holds if the processes are observed up to time T, but this informational inequality fails for longer observation periods; consider e.g. $A(t) = t-1$, $B = 0$, $T = 1$. On the other hand, noncomparable regression models $E(K_1)$ and $E(K_2)$, where $I = [0,T]$ and Θ is dense in $H(K_1)$, cannot be made comparable by longer observation periods (provided, of course, that K_1 and K_2 do not depend on T). This is caused by the fact that the ordering $\geq$ may be characterized by the ordering $\leq$ of the covariance functions K_1 and K_2, see Corollary 2. As a nice consequence one obtains that solutions A and B of the differential inequality (3.11) must satisfy $A \leq B$ on $[0,T]$.

(c) For parametric regression models it may happen that noncomparable models become comparable by longer observation periods, in contrast to nonparametric regression models with a large (dense) set of possible regression functions. A corresponding observation in the setting of iid observations can be found in [26]. For instance, let $K(s,t) = \min(s,t)$, $h(t) = t^2$ and $g(t) = t$ on $[0,T]$. By (3.3), $\|h\|_K^2 = 4T^3/3$ and $\|g\|_K^2 = T$ holds. Hence, by (3.5), $E(h,K) \not\geq E(g,K)$ for $0 < T < (3/4)^{1/2}$, but $E(h,K) \geq E(g,K)$ for $T \geq (3/4)^{1/2}$.

EXAMPLE 6. (Ornstein-Uhlenbeck error processes). Let $A = c$, $c \in \mathbb{R}$, and $\Theta = \{\theta \in W_1^2[0,T] : \theta(0) = 0\}$ (or e.g. Θ = space of polynomials on $[0,T]$ vanishing at zero or Θ = linear span of $\{t \to \cos(n\pi t/T) - 1 : n \in \mathbb{N}\}$). Then

$$K_c(s,t) = (2c)^{-1}\exp(cs+ct) - (2c)^{-1}\exp(c|t-s|) \text{ if } c \neq 0$$

and by Theorem 6, $E(K_c) \geq E(K_d)$, $c,d \in \mathbb{R}$, if and only if $c = d$ in case $c \geq 0$ and $d \geq 0$; $|c| \geq d$ in case $c < 0$ and $d \geq 0$; $c \leq d$ in case $c < 0$ and $d < 0$. This example shows that the "ergodic" case $c,d < 0$ corresponds to comparable experiments while the "nonergodic" case, $c,d > 0$ yields noncomparable experiments. For these notions we refer to Basawa and Scott [2].

EXAMPLE 7. (Brownian bridges). For $c \in \mathbb{R}$, $c \neq 0$, let $A_c(t) = (t-c)^{-1}$ and assume $T < c$ if $c > 0$. Choose Θ as in the preceding example. Then $K_{A_c}(s,t) = \min(s,t) - st/c$ and $E(K_{A_c}) \geq E(K_{A_d})$ if and only if $c \leq d$ in case $c,d > 0$, $T < \min(c,d)$, and in case $c,d < 0$; c,d arbitrary in case $c > 0$, $d < 0$ and $T < c$.

Semiparametric regression models. Let

$$X(t) = \sum_{i=1}^{r} \beta_i h_i(t) + \gamma(t) + Z(t), \qquad t \in I, \tag{3.14}$$

where $h = (h_1, \cdots, h_r) \in H(K)^r$ and γ is assumed to belong to some linear subspace Γ of $H(K)$. Here $\theta = (\beta, \gamma)$, $\Theta = \mathbb{R}^r \times \Gamma$ and $S(\beta, \gamma) = S_h(\beta, \gamma) = \sum_{i=1}^r \beta_i h_i + \gamma$. Given $h \in H(K_1)^r$ and $g \in H(K_2)^r$ ($I_1 = I_2$, $\Gamma \subset H(K_1) \cap H(K_2)$), it follows from Proposition 4 that $E(S_h, K_1) \geq E(S_g, K_2)$ if and only if

$$\begin{aligned} &\sup_{\beta \in \mathbb{R}^r} (2\beta^t f(\gamma) - \beta^t J \beta) \leq \|\gamma\|_{K_1}^2 - \|\gamma\|_{K_2}^2 \text{ for every } \gamma \in \Gamma \text{ with} \\ &f(\gamma) = (\langle h_1, \gamma \rangle_{K_1} - \langle g_1, \gamma \rangle_{K_2}, \cdots, \langle h_r, \gamma \rangle_{K_1} - \langle g_r, \gamma \rangle_{K_2})^t \\ &\quad \text{and } J = J(h, K_1) - J(g, K_2). \end{aligned} \tag{3.15}$$

Since $\sup_{\beta \in \mathbb{R}^r}(2\beta^t f(\gamma) - \beta^t J \beta) = f(\gamma)^t \overline{J} f(\gamma)$ if $f(\gamma) \in J(\mathbb{R}^r)$ and $= \infty$ otherwise, provided $J \geq 0$ (cf. [21], p. 108), this condition is equivalent to

$$J \geq 0, f(\Gamma) \subset J(\mathbb{R}^r) \text{ and } f(\gamma)^t \overline{J} f(\gamma) \leq \|\gamma\|_{K_1}^2 - \|\gamma\|_{K_2}^2 \text{ for every } \gamma \in \Gamma. \tag{3.16}$$

Recall that $\overline{J}$ denotes an arbitrary generalized inverse of J. For semiparametric regression models which differ only by their parametric part, that is, $K_1 = K_2 = K$, (3.16) takes the particular simple form

$$\begin{aligned} &J(h, K) \geq J(g, K) \text{ and } h_i - g_i \\ &\text{belongs to the orthogonal complement of } \Gamma \text{ for every } i = 1, \cdots, r. \end{aligned} \tag{3.17}$$

4. Nonlinear Parametric Models. This section is concerned with the local comparison of smooth regression models (1.1) when Θ is an open subset of $\mathbb{R}^r$ or a (not necessarily open, nondegenerate) subinterval of $\mathbb{R}$. We refer the reader to Ibragimov and Has'minskii [6] and Kutoyants [12] for the estimation theory of such parametric models; for the testing theory in case $r = 1$ see [17].

The differentiability condition involved in the definition of the local ordering of experiments is ensured for differentiable functions S. Here $S : \Theta \to H(K)$ is differentiable at $\theta_0 \in \Theta$ with derivative $\dot{S}(\theta_o) \in H(K)^r$ if

$$\|S(\theta_o + \delta) - S(\theta_o) - \delta^t \dot{S}(\theta_o)\|_K = o((\delta^t \delta)^{1/2}) \text{ as } \delta \to 0.$$

Then $\dot{S}(\theta_o, t) = \nabla_\theta S(\theta_o, t)$ for every $t \in I$ and $\sup_{t \in I} |S(\theta_o + \delta, t) - S(\theta_o, t) - \delta^t \dot{S}(\theta_o, t)| = o((\delta^t \delta)^{1/2})$ since the inclusion of $H(K)$ into $C(I)$ is continuous.

THEOREM 7. *Let $\theta_o \in \Theta$ and suppose that $S_j : \Theta \to H(K_j)$ is differentiable at θ_o with derivative $\dot{S}_j(\theta_o) \in H(K_j)^r$, $j = 1, 2$. Then $E(S_1, K_1) \geq_{\theta_o} E(S_2, K_2)$ if and only if*

$$J(\dot{S}_1(\theta_o), K_1) \geq J(\dot{S}(\theta_o), K_2). \tag{4.1}$$

Proof. We suppress the index j. Computing the likelihood ratio $f_{\theta_o,\theta} = dP_\theta/dP_{\theta_o}$ yields

$$f_{\theta_o,\theta} = \exp\{L_K(S(\theta) - S(\theta_o)) - \|S(\theta)\|_K^2/2 + \|S(\theta_o)\|_K^2/2\},$$

see (2.2). Since normal random variables have finite exponential moments, $f_{\theta_o,\theta} \in L^2(P_{\theta_o})$ for every $\theta \in \Theta$. We claim that $\Theta \to L^2(P_{\theta_o})$, $\theta \to f_{\theta_o,\theta}$ is differentiable at θ_o with derivative $\dot{f}_{\theta_o} \in L^2(P_{\theta_o})^r$ given by

$$\dot{f}_{\theta_o,i} = L_K(\dot{S}(\theta_o)_i) - \langle \dot{S}(\theta_o)_i, S(\theta_o)\rangle_K, \qquad i = 1, \cdots, r. \tag{4.2}$$

Observe that by (2.4), $E_{\theta_o}\dot{f}_{\theta_o,i} = 0$ and

$$\mathrm{Cov}_{\theta_o}\dot{f}_{\theta_o} = J(\dot{S}(\theta_o), K). \tag{4.3}$$

Furthermore, we have

$$E_{\theta_o}(f_{\theta_o,\theta})^2 = \exp(\|S(\theta) - S(\theta_o)\|_K^2)$$

and

$$E_{\theta_o}(f_{\theta_o,\theta} \cdot \dot{f}_{\theta_o,i}) = E_\theta \dot{f}_{\theta_o,i} = \langle \dot{S}(\theta_o)_i, S(\theta) - S(\theta_o)\rangle_K.$$

This yields

$$\begin{aligned}
&E_{\theta_o}(f_{\theta_o,\theta_o+\delta} - 1 - \delta^t \dot{f}_{\theta_o}^2)^2 \\
&= E_{\theta_o}(f_{\theta_o,\theta_o+\delta})^2 - 1 - 2\, E_{\theta_o}(f_{\theta_o,\theta_o+\delta} \cdot \delta^t \dot{f}_{\theta_o}) + E_{\theta_o}(\delta^t \dot{f}_{\theta_o})^2 \\
&= \exp(\|S(\theta_o+\delta) - S(\theta_o)\|_K^2) - 1 - 2\langle \delta^t \dot{S}(\theta_o), S(\theta_o+\delta) \\
&\qquad - S(\theta_o)\rangle_K + \|\delta^t \dot{S}(\theta_o)\|_K^2 \\
&= \exp(\|S(\theta_o+\delta) - S(\theta_o)\|_K^2) - 1 - \|S(\theta_o+\delta) - S(\theta_o)\|_K^2 \\
&\quad + \|S(\theta_o+\delta) - S(\theta_o) - \delta^t \dot{S}(\theta_o)\|_K^2 \\
&= o(\delta^t\delta) \text{ as } \delta \to 0 \qquad (\delta \in \mathbb{R}^r)
\end{aligned}$$

and our claim is proved. It follows that $E(S, K)$ is $L^1(P_{\theta_o})$-differentiable with the same derivative (cf. [29], Satz 1.199 and Satz 1.190). Now, noting that $\mathcal{L}(\dot{f}_{\theta_o} \mid P_{\theta_0}) = N(o, J(\dot{S}(\theta_o), K))$, the convex function criterion (1.6) yields the assertion. ∎

Equation (4.3) shows that $J(\dot{S}_j(\theta_o), K_j)$ is the Fisher information matrix of $E(S_j, K_j)$ at θ_o. Therefore, the local ordering of regression models with differentiable functions S coincides with the ordering $\geq$ of the Fisher information matrices. In case $r = 1$, the local ordering $\geq_{\theta_o}$ of arbitrary $L^1(P_{\theta_o})$-differentiable experiments may be completely described by the slopes of power functions of the locally most powerful level α tests for testing $\theta = \theta_o$ vs. $\theta > \theta_o$

(cf. [28]). This is obvious in our setting, since we have in $E(S,K)$ by (4.2) and the generalized Neyman-Pearson lemma

$$\sup\{E_{\theta_o}(\varphi \dot{f}_{\theta_o}) : E_{\theta_o}\varphi = \alpha,\ \ \varphi \text{ test}\} = \|\dot{S}(\theta_o)\|_K \int_{z_\alpha}^{\infty} y\, dN(0,1)(y)$$

for every $\alpha \in (0,1)$.

EXAMPLE 8. (Wiener processes with a nonlinear drift). Let $\Theta = [0,\infty)$, $S_1(\theta,t) = \log(\theta t+1)$, $S_2(\theta,t) = (\theta t+1)^{1/2}-1$ and $K_1 = K_2 = K$ with $K(s,t) = \min(s,t)$ on $[0,T]$. Then $\dot{S}_1(0,t) = t$ and $\dot{S}_2(0,t) = t/2$ are the $H(K)$-derivative of S_1, S_2 at 0, respectively, and $\|\dot{S}_1(0)\|_K^2 = T$ and $\|\dot{S}_2(0)\|_K^2 = T/4$, see (3.3). Hence, $E(S_1,K)$ is locally more informative than $E(S_2,K)$ at $\theta_o = 0$. Clearly, $E(S_1,K)$ is not (globally) more informative than $E(S_2,K)$ by Theorem 1, since e.g. $R_{1,0}(1,1) = T/(T+1)$, $R_{2,0}(1,1) = \log(T+1)/4$ and hence $R_{1,0}(1,1) < R_{2,0}(1,1)$ at least for $T \geq e^4 - 1$.

Theorem 7 may be applied to the local comparison of sampling designs along the lines described in Section 3.

5. Information Contained in Additional Observation Periods. For the discussion of this topic, consider the parametric regression model

$$X(t) = \sum_{i=1}^{r} \theta_i h_i(t) + Z(t), \qquad t \geq 0, \tag{5.1}$$

where $\Theta = \mathbb{R}^r$ and $Z(t)$, $t \geq 0$, is a mean-zero Gaussian process with continuous sample paths and covariance function K. The restriction h_T of $h = (h_1,\cdots,h_r)$ on $[0,T]$ is assumed to belong to the r-fold product of the RKHS of the restriction K_T of K on $[0,T]\times[0,T]$ and $J(h,K)_T$ is assumed to be regular for every $T > 0$, where $J(h,K)_T$ is short for $J(h_T,K_T)$. Observation of X up to time T corresponds to the experiment $E(h_T,K_T)$ which is simply denoted by E_T.

Let $f : (0,\infty) \to (0,\infty)$ be a function satisfying $f(t) > t$ for every $t \in (0,\infty)$. Clearly, $E_{f(T)}$ is at least as informative as E_T. How large is the amount of information contained in the additional observation period $[T, f(T)]$? This quantity is measured in terms of the deficiency of E_T relative to $E_{f(T)}$. Since $(L_{K_T}(h_1),\cdots,L_{K_T}(h_r))$ is a sufficient statistic for E_T, see (2.2), we find that

$$E_T \sim (N(\theta, J(h,K)_T^{-1}) : \theta \in \mathbb{R}^r).$$

Therefore, according to a result of Torgersen [25] (cf. also [13], p. 130, [24]) on the comparison of translation experiments, we obtain

$$\delta(E_T, E_{f(T)}) = \|N(0, J(h,K)_T^{-1}) - N(0, J(h,K)_{f(T)}^{-1})\|. \tag{5.2}$$

This is a rather complicated function of the Fisher information matrices of E_T and $E_{F(T)}$. So the following bounds may be helpful. They are the usual bounds for the variational distance using the Hellinger distance (cf. [23], 2.15) and for the above normal distributions given by

$$2\{1-[\det(J_T J_{f(T)})]^{1/4}[\det((J_T+J_{f(T)})/2)]^{-1/2}\} \le \delta(E_T, E_{f(T)}) \\ \le 2\{1-[\det(J_T J_{f(T)})]^{1/2}[\det((J_T+J_{f(T)})/2)]^{-1}\}^{1/2}$$

with $J_T = J(h,K)_T$. An immediate consequence is that $\delta(E_T, E_{f(T)}) \to 0$ if $J_T^{-1} J_{f(T)} \to I_r$ as $T \to \infty$.

Now let $r = 1$ and $h = h_1$. Let$\|h\|_T$ be short for $\|h_T\|_{K_T}$. Then (5.2) yields

$$\delta(E_T, E_{f(T)}) = G(\|h\|^2_{f(T)}/\|h\|^2_T), \quad \text{where} \tag{5.3}$$
$$G(x) = 4\left\{\Phi\left(\left(\frac{x\log x}{x-1}\right)^{1/2}\right) - \Phi\left(\left(\frac{\log x}{x-1}\right)^{1/2}\right)\right\} \quad \text{if } x>1 \text{ and } G(1)=0.$$

Observe that $G : [1,\infty) \to [0,2]$ is continuously differentiable with bounded derivative and strictly increasing. Setting $C = \sup\{G'(x) : x \ge 1\}$, we get for every $x \ge 1, T > 0$

$$|\delta(E_T, E_{f(T)}) - G(x)| \le C|\ \|h\|^2_{f(T)}/\|h\|^2_T - x|.$$

The most interesting case is

$$\lim_{T\to\infty} \|h\|^2_{f(T)}/\|h\|^2_T = 1. \tag{5.4}$$

Then, by (5.3) and Taylor expansion of G about 1, we obtain

$$\delta(E_T, E_{f(T)}) = (2/\pi e)^{1/2}(\|h\|^2_{f(T)}/\|h\|^2_T - 1) \\ + o(\|h\|^2_{f(T)}/\|h\|^2_T - 1) \text{ as } T \to \infty \tag{5.5}$$

since $G'(1) = (2/\pi e)^{1/2}$.

EXAMPLE 9. Let $K(s,t) = \min(s,t)$, $h(t) = t^n$, $n \ge 1$, and $f(T) = T + o(T)$ as $T \to \infty$. Then by (3.3), $\|h\|^2_T = n^2T^{2n-1}/(2n-1)$ and hence $\delta(E_T, E_{f(T)}) = (2/\pi e)^{1/2}((f(T)/T)^{2n-1} - 1) + o((f(T)/T)^{2n-1} - 1)$, $T \to \infty$. In particular, if $h(t) = t$ and $f(T) = T + c$ for some $c > 0$, we obtain $\delta(E_T, E_{f(T)}) = (2/\pi e)^{1/2}cT^{-1} + o(T^{-1})$, for $f(T) = T + T^{1/2}$ we get $\delta(E_T, E_{f(T)}) = (2/\pi e)^{1/2}T^{-1/2} + o(T^{-1/2})$, and for $f(T) = T + \log T$, $\delta(E_T, E_{f(T)}) = (2/\pi e)^{1/2}T^{-1}\log T + o(T^{-1}\log T)$.

REFERENCES

[1] Bandemer, H., Näther, W. and Pilz, J. (1987). Once more: Optimal experimental design for regression models. *Statistics* **18**, 171–217.

[2] Basawa, I. V. and Scott, D. J. (1983). *Asymptotic Optimal Inference for Non-ergodic Models.* Lecture Notes in Statistics **17**, Springer, New York.

[3] Hansen, O. H. and Torgersen, E. N. (1974). Comparison of linear normal experiments. *Ann. Statist.* **2**, 367–373.

[4] Heyer, H. (1982). *Theory of Statistical Experiments.* Springer, New York.

[5] Heyer, H. (1988). Bemerkungen zum Vergleich linearer normaler Experimente. *Monatsh. Math.* **106**, 9–23.

[6] Ibragimov, I. A. and Has'minskii, R. Z. (1981). *Statistical Estimation.* Springer, New York.

[7] Ibragimov, I. A. and Khas'minskii, R. Z. (1984). On nonparametric estimation of the value of a linear functional in Gaussian white noise. *Theory Probab. Appl.* **29**, 18–32.

[8] Ibragimov, I. A., Nemirovskii, A. S. and Khas'minskii, R. Z. (1986). Some problems on nonparametric estimation in Gaussian white noise. *Theory Probab. Appl.* **31**, 391–406.

[9] Janssen, A. (1988). A convolution theorem for the comparison of exponential families. Technical Report 211, University of Siegen.

[10] Jørsboe, O. G. (1968). Equivalence or singularity of Gaussian measures on function spaces. Various Publications Series No. 4, University of Aarhus.

[11] Kuo, H. H. (1975). *Gaussian Measures in Banach Spaces.* Lecture Notes in Math. **463**, Springer, New York.

[12] Kutoyants, Y. A. (1984). *Parameter Estimation for Stochastic Processes.* Helderman Verlag, Berlin.

[13] LeCam, L. (1986). *Asymptotic Methods in Statistical Decision Theory.* Springer, New York.

[14] Lipster, R. S. and Shiryayev, A. N. (1977). *Statistics of Random Processes* I. Springer, New York.

[15] Luschgy, H. (1987). Comparison of shift experiments on a Banach space. In M. L. Puri et al. (Eds.), *Mathematical Statistics and Probability Theory*, Vol. A, 217–230. Reidel, Dordrecht.

[16] Luschgy, H. (1989). Comparison of location models for stochastic processes, mimeo.

[17] LUSCHGY, H. (1991). Testing one-sided hypotheses for the mean of a Gaussian process. *Metrika* **38**.

[18] MILBRODT, H. and STRASSER, H. (1985). Limits of triangular arrays of experiments. In A. Janssen et al., *Infinitely Divisible Statistical Experiments*, 14–54. Lecture Notes in Statist. **27**, Springer, New York.

[19] NÄTHER, W. (1985). Effective observation of random fields. *Teubner-Texte Math.* **72**, Teubner Verlag, Leipzig.

[20] PARZEN, E. (1967). *Time Series Analysis Papers.* Holden-Day, San Francisco.

[21] ROCKAFELLAR, R. T. (1970). *Convex Analysis.* Princeton University Press, Princeton, NJ.

[22] STEPNIAK, C. and TORGERSEN, E. N. (1981). Comparison of linear models with partially known covariances with respect to unbiased estimation. *Scand. J. Statist.* **8**, 183–184.

[23] STRASSER, H. (1985). *Mathematical Theory of Statistics.* De Gruyter, Berlin, New York.

[24] SWENSEN, A. R. (1980). Deficiencies between linear normal experiments. *Ann. Statist.* **8**, 1142–1155.

[25] TORGERSEN, E. N. (1972). Comparison of translation experiments. *Ann. Math. Statist.* **43**, 1383–1399.

[26] TORGERSEN, E. N. (1976). Comparison of statistical experiments. *Scand. J. Statist.* **3**, 186–208.

[27] TORGERSEN, E. N. (1984). Orderings of linear models. *J. Statist. Plann. Inference* **9**, 1–17.

[28] TORGERSEN, E. N. (1985). Majorization and approximate majorization for families of measures, applications to local comparison of experiments and the theory of majorization of vectors in R^n (Schur convexity). In T. Calinski and W. Klonecki (Eds.), *Linear Statistical Inference*, pp. 259–310. Lecture Notes in Statist. **35**, Springer, New York.

[29] WITTING, H. (1985). *Mathematische Statistik I.* Teubner, Stuttgart.

INSTITUT FÜR MATHEMATISCHE STATISTIK
UNIVERSITÄT MÜNSTER
EINSTEINSTRASSE 62
D-4400 MÜNSTER
FEDERAL REPUBLIC OF GERMANY