The Annals of Statistics
1993, Vol. 21, No. 2, 943–964

CONTRASTS UNDER LONG-RANGE CORRELATIONS

BY H. KÜNSCH, J. BERAN, AND F. HAMPEL

ETH Zürich, Texas A & M University and ETH Zürich

The background of the paper is the empirical observation from a variety of subject areas that long-range correlations appear to be much more frequent than has been previously assumed. This includes high-quality measurement series which are commonly treated as prototypes of "i.i.d." observations. Evidence is briefly cited in the paper. It has already been shown elsewhere that long-range dependence leads to results that can be qualitatively different from those obtained under short-range dependence, and in particular, that long-range dependence has drastic effects on the naive statistical treatment of *absolute constants*. The natural question arising from this, also of relevance for statistical practice, is how the long-range dependence affects the statistics for *contrasts*. The main answer given in this paper is twofold.

(i) If the experimental conditions are well mixed as provided by randomization, the *levels* of tests and confidence intervals derived under the independence assumption are still *correct*, asymptotically and usually in good approximation for finite samples.

(ii) Even under randomly mixed designs there are typically large unnoticed *power* and *efficiency losses* due to the long-range dependence. They can be greatly reduced without estimating the correlations by a simple blocking device.

1. Introduction. This paper grew out of an attempt to understand why some parts of statistical applications are successful (according to folklore in applications), despite the surprisingly widespread occurrence of long-range correlations (see below) which are not considered in the analysis. The answer, but briefly, is that the principal effects of long-range correlations, so disastrous for absolute constants, cancel out for contrasts if the treatments are applied in random order. These results provide a partial justification for a major part of regression and analysis of variance methodology, in which the data normally are treated as if they were independent. (Intercepts and grand means are another matter.) On the other hand, it turned out that with the usual methods for linear models there are efficiency and power losses due to long-range correlations which can be considerable. These losses can be partly avoided not only by cumbersome estimation and explicit inclusion of the long-range correlations into the inference process, but also simply by the use of randomized block designs (again corroborating a good applied practice from another theoretical angle).

Received October 1990; revised June 1992.

AMS 1991 *subject classifications*. Primary 62M10; secondary 62F35, 62J10, 62M99.

Key words and phrases. Long-range dependence, contrasts, ordinary and generalized least squares, regression with correlated errors, robustness of validity, robustness of efficiency, randomization, blocking.

Put differently, the levels of tests and so on for effects and regression slopes are still approximately correct if long-range correlated data are treated as independent, as is frequently the case. This causes statisticians to make correct positive claims on effects with the controlled error probability of the first kind; but there are—partly considerable—unnoticed power and efficiency losses, causing statisticians to discover too few true effects.

This situation is markedly different from that for absolute constants (grand means, intercepts) where the effects of long-range correlations are disastrous already for the level even in moderately small samples. This is probably a major reason for the disdain many physicists have towards statistics for their constants [cf. Jeffreys (1939), see pages 301 and 270–271 of the 3rd edition], and it turns out to be a remarkably wise custom not to include the grand mean into the analysis of variance table. For confidence and prediction intervals for the mean, one really has to take the correlations into account ["one-sample t-test for long-range dependence," see Beran (1986), (1989)].

For simplicity we restrict ourselves to serial correlations assuming time ordered observations, although more general structures can also be of interest. The standard models for serial dependence are ARMA-processes. They are well suited for describing an arbitrary behavior of the small lag correlations, but asymptotically their correlations decay exponentially fast and their spectrum is bounded and continuous. By standard results this implies that the variance of the arithmetic mean eventually decays like n^{-1}. Such properties are in striking contrast to the empirical behavior of a large number of observed time series from all subject areas including even high quality measurement series. There the spectrum seems to be unbounded at the origin and the variance of the arithmetic mean decays like $n^{-\alpha}$ with some $\alpha < 1$ even for quite large n's. Stationary processes whose correlations decay so slowly that their sum is infinite can model such a behavior. This is what we mean by long-range correlations. Since not all readers may be familiar with this topic and some may doubt its practical importance, we cite in Section 2 some literature which collects empirical evidence for long-range correlation. In view of this evidence it seems crucial to us to work with long-range correlated errors although this imposes some mathematical difficulties, for instance Rosenblatt's strong mixing property does not hold. For a general survey of statistical methods for models with long-range correlations see Künsch (1987) and Beran (1992).

In some areas of applications, in particular for field experiments, it is obvious that the assumption of independent errors is not tenable. In order to avoid such an assumption, Fisher (1925) introduced the techniques of randomization and blocking. Our results provide a justification of these techniques different from the one usually given in the literature. Results on the validity of randomization [see, e.g., Bailey and Rowley (1987)] take the plot effect to be fixed and consider the distribution induced by the randomization. Here we argue conditional on the treatment allocation and consider a standard linear model with correlated errors. We show that with high probability randomization chooses a design matrix for which the usual i.i.d. inference about contrasts is approximately correct. Moreover we show that blocking greatly reduces the variance of contrasts even though in our model different blocks are not independent and the block effect is not constant within a block. We also compute the efficiency losses due to ignoring correlations: They are large without blocking, but negligible with blocking.

There is already a considerable amount of mathematical results on design and analysis of experiments with serially correlated error structures. The special role played by contrasts has been realized and attention was mostly restricted to these. The difference between these papers and our work is two-fold. First the correlation structures in the earlier work allowed only short term correlations [e.g., Williams (1952), Jenkins and Chanmugan (1962), Berenblut and Webb (1974), Kiefer and Wynn (1981), (1984), Cheng (1983), Martin (1986) and Morgan and Chakravarti (1988)]. Second the goal of many of these papers is to construct in some sense optimal designs for prespecified correlation structures. By contrast, we consider a broad and realistic class of correlation structures which are close to a nonstationary behavior, and we investigate what actually happens to common statistics, tests and confidence intervals with designs which are in good approximation frequently used. In other words, no attempt is made in this paper to estimate the long-range correlations or to fit a specific model with long-range correlations to the data. This would be impractical or even impossible in many situations.

The problems considered here belong to the area of regression with correlated errors. In the case of short-range correlations some of our results can be derived from the classical theorems of Grenander and Rosenblatt [(1957), Section 7.3]. Still we have not seen these results stated explicitly in the literature despite their importance for statistical practice. Moreover, for long-range dependence the theorems of Grenander and Rosenblatt do no longer apply since the spectrum has a singularity at zero. So far no general theory for regression with long-range correlated errors has been developed. Yajima [(1988), (1991)] covers the case of deterministic regressors, but it is an open problem to verify his conditions in our case. We use here the special structure of the regressors to construct a martingale and then apply the martingale convergence theorem.

Our main results are asymptotic, but in a situation where dependence does not become asymptotically negligible. Moreover, we give also results on the speed of convergence and on the bias of the estimated variance. Finally, in Section 7 we give some simulation studies which complement our theoretical results.

2. Evidence for long-range correlations. We believe that for applied statisticians it should be vital to know the full range of applicability of a stochastic model. But due to limitation of space we shall say only a few words about the occurrence of long-range correlations in real data. This seems the more defensible as there are already two surveys on this topic [Hampel, Ronchetti, Rousseeuw and Stahel (1986), Chapter 8.1 and Hampel (1987)] to which the reader is referred for more details and references. Areas of application discussed and cited there include economics, ecology, quality control, sample surveys, linguistics and even measurement series in the "hard sciences."

For geophysical data, Mandelbrot [cf., e.g., Mandelbrot and Wallis (1968), (1969)] showed the widespread occurrence of long-range dependence and made the model of self-similar processes [Kolmogorov (1940)] popular before he integrated it into his treatment of fractals. He stirred up a considerable controversy in hydrology because many scientists did not want to give up their traditional models of short-range dependence.

In electrical engineering and, more generally, the frequency analysis of time series, "$1/f$ noises" and the "infrared catastrophy," referring to a pole of the spectrum at the origin, showed up increasingly in a broad variety of examples [cf., e.g., Voss (1979) and Percival (1985)].

Perhaps most unbelievable to many is the observation that high-quality measurement series from astronomy, physics and chemistry, generally considered as prototypes for "i.i.d." observations, are not independent but long-range correlated, as can now be proven by the test of Graf (1983); see Graf, Hampel and Tacier (1984). This fact was qualitatively known to and stressed by eminent statisticians such as Newcomb (1895), Pearson (1902), Student (1927) and Jeffreys (1939), but was apparently ignored or suppressed for various reasons.

In the light of these observations, there is some interest in understanding the effects of long-range correlations on level and on power of methods for contrasts derived under the independence assumption, and on simple improvements for them, as investigated in the following parts of this paper.

3. Notations and definitions.

3.1. *Contrasts.* We consider in this paper a one-way layout with p treatments. Let β_j denote the mean under the jth treatment. A linear combination of treatment means $\sum_{j=1}^{p} \lambda_j \beta_j$ with $\sum_{j=1}^{p} \lambda_j = 0$ is called a *contrast*. Obviously the set of all contrasts forms a $(p-1)$-dimensional vector space. Multiway layouts can be reduced to this case in the following way. If we have k factors on $p_1, \dots, p_k$ levels, respectively, we put $p = \Pi p_j$ and take each possible level combination as one treatment. All main effects and interactions in the original layout become then contrasts in the one-way layout. In our asymptotics the number of treatments p will be fixed whereas the number of observations increases to infinity. If we want to include fractional factorials, we have to let p increase at the same speed as n. This is a problem for further research.

3.2. *Randomization.* The p treatments are applied sequentially in time or along a transect in the plane. Denote the treatment allocation at time t by $a_t \in \{1, 2, \dots, p\}$. Hence the design matrix $X_n = (x_{t,j};\ 1 \le t \le n;\ 1 \le j \le p)$ is given by

$$x_{t,j} = \begin{cases} 1, & \text{if } a_t = j, \\ 0, & \text{otherwise} \end{cases} \tag{3.1}$$

and the model becomes

$$y_t = \sum_{j=1}^{p} \beta_j x_{t,j} + \varepsilon_t. \tag{3.2}$$

The treatment allocations a_t are supposed to be random. We will consider three different randomization schemes. The first one is *complete randomization* which means that

$$(a_t) \text{ is i.i.d. with } P[a_t = j] = \pi_j, \qquad j = 1, \dots, p. \tag{3.3}$$

This makes the number of observations with a given treatment, $n_j = \sum_{t=1}^{n} x_{t,j}$, $j = 1, \dots, p$, random. With *restricted randomization* we fix n_j in advance and use the conditional distribution given the n_j's:

(3.4) All $n!/(n_1! \cdots n_p!)$ allocations $(a_1, \dots, a_n)$ with $\sum_{t=1}^{n} x_{t,j} = n_j$, $j = 1, \dots, p$, are equally likely.

For large n the difference between (3.3) and (3.4) with $n_j = n\pi_j$ will be small. The first one is somewhat easier to analyze because we can keep the allocations already chosen as n increases.

The third randomization scheme to be considered is *blockwise randomization*. Here we form blocks $B_k = \{(k-1)l + 1, (k-1)l + 2, \dots, kl\}$ of length l, $k = 1, \dots, b = n/l$. For simplicity we consider only complete blocks. Thus we require:

(3.5) The allocations in different blocks are independent and in each block B_k all allocations with $\sum_{t \in B_k} x_{t,j} = l_j \geq 1$, $j = 1, \dots, p$, are equally likely.

Obviously $\sum_{j=1}^{p} l_j = l$. The most common case is $l = p$, $l_j = 1$ for all j, that is, each treatment occurs exactly once in each block.

3.3. *Correlations of the errors.* The errors ε_t in the model (3.2) are supposed to have expectation zero, constant variance σ^2 and long-range correlations. We assume that either

$$|\rho(t,s)| = |\mathrm{Corr}(\varepsilon_t, \varepsilon_s)| \leq \text{const.}|t-s|^{2H-2} \qquad (t \neq s) \text{ with } H \in \left(\tfrac{1}{2}, 1\right) \tag{3.6}$$

or in the stationary case $\rho(t,s) = \rho(t-s)$ with

$$\begin{gathered} \rho(t-s) = \int_{-\pi}^{\pi} e^{i\omega(t-s)} f(\omega)\, d\omega, \quad \text{where } \int_{-\pi}^{\pi} f(\omega)\, d\omega = 1, \\ f(\omega) = |\omega|^{1-2H} f_0(\omega) \quad \text{with } H \in \left(\tfrac{1}{2}, 1\right) \end{gathered} \tag{3.7}$$

and $f_0(\omega)$ continuous, of bounded variation and bounded away from zero. Condition (3.7) implies (3.6), see Zygmund [(1959), Chapter V.2]. In (3.6), stationarity is not required. The two most common models with long-range correlations, fractional noise [Mandelbrot and van Ness (1968)] and fractional ARIMA (p, d, q)-processes [Hosking (1981)], satisfy (3.7), see Sinai (1976) in the former and Hosking (1981) in the latter case.

If the treatments are allocated in space as for instance in agriculture, the index t becomes multidimensional, that is, the ε_t's are then a random field. We expect our results to generalize to this case.

3.4. *Estimators.* The ordinary least squares estimator (OLSE) in the models (3.1) and (3.2) is

$$(3.8)\quad \hat{\beta}(n) = \left(X_n^T X_n\right)^{-1}\left(X_n^T y\right) = \left(\sum_t y_t x_{t,1}/n_1, \dots, \sum_t y_t x_{t,p}/n_p\right)^T.$$

Here we are interested in the variance of a contrast $\sum \lambda_j \hat{\beta}_j(n)$ given the design X_n because the design is ancillary. It is sufficient to consider the variances $V_n(j, k|X_n)$ of $\hat{\beta}_j(n) - \hat{\beta}_k(n)$ because $\sum \lambda_j \hat{\beta}_j(n) = \sum \lambda_j(\hat{\beta}_j(n) - \hat{\beta}_1(n))$ and $\operatorname{Cov}(\hat{\beta}_j(n) - \hat{\beta}_1(n), \hat{\beta}_k(n) - \hat{\beta}_1(n)) = (1/2)(\operatorname{Var}(\hat{\beta}_j(n) - \hat{\beta}_1(n)) + \operatorname{Var}(\hat{\beta}_k(n) - \hat{\beta}_1(n)) - \operatorname{Var}(\hat{\beta}_j(n) - \hat{\beta}_k(n)))$. From (3.8) it follows immediately that

$$(3.9)\quad V_n(j, k|X_n) = \sigma^2 \sum_{t,s=1}^{n} \left(x_{t,j}/n_j - x_{t,k}/n_k\right)\rho(t,s)\left(x_{s,j}/n_j - x_{s,k}/n_k\right).$$

If the ε_t's were uncorrelated, $V_n(j, k|X_n)$ would be $\sigma^2(n_j^{-1} + n_k^{-1})$. Hence if one is not aware of the correlations, one estimates $V_n(j, k|X_n)$ by $\hat{\sigma}^2(n_j^{-1} + n_k^{-1})$, where

$$(3.10)\quad \begin{aligned}\hat{\sigma}^2 &= (n-p)^{-1} \sum_{t=1}^{n} \left(y_t - \sum_t \hat{\beta}_j(n) x_{t,j}\right)^2 \\ &= (n-p)^{-1}\left(\sum_t \varepsilon_t^2 - \sum_j n_j^{-1}\left(\sum_t \varepsilon_t x_{t,j}\right)^2\right).\end{aligned}$$

In Section 4 we are investigating the difference $\hat{\sigma}^2(n_j^{-1} + n_k^{-1}) - V_n(j, k|X_n)$ under the assumptions (3.4) and (3.6). This is relevant for the question of robustness of the level of tests and confidence intervals (robustness of validity) under deviations from independence. In Section 5 the same question is investigated under assumptions (3.3) and (3.7).

If we are interested in the robustness of power (robustness of efficiency) we have to study the efficiency of the ordinary least squares estimator. In Section 5 we will compare $V_n(j, k|X_n)$ with the conditional variance of the best linear unbiased estimator (BLUE) under assumptions (3.3) and (3.7). The best linear unbiased estimator $\tilde{\beta}(n)$ is given by

$$(3.11)\quad \tilde{\beta}(n) = \left(X_n^T \Sigma_n^{-1} X_n\right)^{-1}\left(X_n^T \Sigma_n^{-1} y\right),$$

where Σ_n is the covariance matrix of $(\varepsilon_1, \dots, \varepsilon_n)$. The variance of $\tilde{\beta}_j(n) - \tilde{\beta}_k(n)$ given the design X_n is then

$$(3.12)\quad V_n^*(j, k|X_n) = \operatorname{Var}\left[\tilde{\beta}_j(n) - \tilde{\beta}_k(n)|X_n\right] = \lambda^T\left(X_n^T \Sigma_n^{-1} X_n\right)^{-1}\lambda$$

with $\lambda = (\lambda_1, \dots, \lambda_p)^T$ such that $\sum \lambda_i \beta_i = \beta_j - \beta_k$.

Finally in Section 6 we will compare $V_n(j,k|X_n)$ under the two different randomizations (3.3) and (3.5), that is, we will see how much we gain by changing the randomization instead of changing the method of estimation. With randomized blocks one will take the estimated block effects $\hat{\mu}_k = l^{-1}\Sigma_{t\in B_k}(y_t - \Sigma_j\hat{\beta}_j x_{t,j})$ into account for the estimation of the variance of the errors. We thus consider instead of (3.10)

$$\hat{\sigma}^2_{\text{block}} = (n-p-b+1)^{-1}\sum_{t=1}^{n}\Big(y_t - \sum_j \hat{\beta}_j x_{t,j} - \sum_k \hat{\mu}_k 1_{[t\in B_k]}\Big)^2$$

$$(3.13)\qquad = (n-p-b+1)^{-1}\Bigg(\sum_t \varepsilon_t^2 - \sum_j n_j^{-1}\Big(\sum_t \varepsilon_t x_{t,j}\Big)^2 - l^{-1}\sum_k\Big(\sum_{t\in B_k}\varepsilon_t\Big)^2 + n^{-1}\Big(\sum_t \varepsilon_t\Big)^2\Bigg).$$

4. Restricted randomization with ordinary least squares.

4.1. *Asymptotic behavior of the variance.* The variance $V_n(j,k|X_n)$ of a contrast depends on the random design matrix X_n and is thus a random variable. Its expectation and variance under the assumption (3.4) is given in the following:

PROPOSITION 4.1.

(i) $E[V_n(j,k|X_n)] = (n_j^{-1} + n_k^{-1})\sigma^2(1 - n^{-1}(n-1)^{-1}\Sigma_{t\neq s}\rho(t,s))$.

(ii) *If* $n_j/n \to \pi_j \in (0,1)$, $j = 1,\dots,p$, *then* $\operatorname{Var}[V_n(j,k|X_n)] \sim (n_j^{-1} + n_k^{-1})^2\sigma^4(2n^{-2}\Sigma'\rho(t,s)^2 - 4n^{-3}\Sigma'\rho(t,s)\rho(s,u) + 2n^{-4}\Sigma'\rho(t,s)\rho(u,v))$, *where* Σ' *denotes summation over indices which are all different.*

The proof is given in the Appendix. From this we can obtain easily the asymptotic behavior of $V_n(j,k|X_n)$.

THEOREM 4.1. *If* $n_j/n \to \pi_j \in (0,1)$, $j = 1,\dots,p$, *and* $\rho(t,s)$ *satisfies* (3.6), *then*

$$E\Big[n\Big(V_n(j,k|X_n) - \big(n_j^{-1} + n_k^{-1}\big)\sigma^2\Big)\Big] = O(n^{2H-2}),$$

$$\operatorname{Var}\big[nV_n(j,k|X_n)\big] = O(c_n)$$

where

$$c_n = n^{-1}, \qquad H < \tfrac{3}{4},$$

$$c_n = \log(n)/n, \qquad H = \tfrac{3}{4},$$

$$c_n = n^{4H-4}, \qquad H > \tfrac{3}{4}.$$

In particular by Chebyshev's inequality $nV_n(j, k|X_n) \to_P (\pi_j^{-1} + \pi_k^{-1})\sigma^2$. The proof is again given in the Appendix. This says that for n large randomization will pick with high probability a design such that the variance of a contrast is almost the same as in the i.i.d. case. The correlations of the errors do however affect the speed of convergence in this asymptotic result: The larger H, the slower it is. Moreover for $H > 3/4$ the bias is of the same order as the standard deviation.

4.2. *Asymptotic behavior of* $\hat{\sigma}^2$. First we show that $\hat{\sigma}^2$ is consistent. The effect of using the residuals instead of the true errors is by (3.10) $-(n-p)^{-1}\Sigma_j n_j^{-1}(\Sigma_t \varepsilon_t x_{t,j})^2$. The following lemma shows that this is asymptotically negligible.

LEMMA 4.1. *If* (3.6) *holds,* $E[(n-p)^{-1}\Sigma_j n_j^{-1}(\Sigma_t \varepsilon_t x_{t,j})^2 | X_n] = O(n^{2H-2})$ *uniformly for all* X_n.

PROOF. See the Appendix.

By the Chebyshev inequality this implies that $(n-p)^{-1}\Sigma n_j^{-1}(\Sigma \varepsilon_t x_{t,j})^2$ is $O_p(n^{2H-2})$ both conditionally on the design X_n and unconditionally. As a direct consequence we have the following:

THEOREM 4.2. *If* (ε_t) *is stationary and ergodic,* $\hat{\sigma}^2 \to_P \sigma^2$ *both conditionally on the design matrix and unconditionally. Hence* $n(V_n(j, k|X_n) - \hat{\sigma}^2(n_j^{-1} + n_k^{-1})) \to_P 0$, *that is, asymptotically the correlations of the errors can be neglected when estimating the variance of a contrast.*

Let us look at finite sample sizes. By (3.10) we obtain the conditional expectation of $\hat{\sigma}^2$ given the design matrix

$$(4.1) \qquad E\left[\hat{\sigma}^2 | X_n\right] = \sigma^2\left(1 - (n-p)^{-1} \sum_{t \neq s} \rho(t,s) \sum_j n_j^{-1} x_{t,j} x_{s,j}\right).$$

Hence the correlated errors make $\hat{\sigma}^2(n_j^{-1} + n_k^{-1})$ a biased estimator of $V_n(j,k|X_n)$. Still on the average we have unbiasedness

$$\left(n_j^{-1} + n_k^{-1}\right) E\left[E\left[\hat{\sigma}^2 | X_n\right]\right] = E\left[V_n(j,k|X_n)\right].$$

This follows from validity of randomization [see, e.g., Bailey and Rowley (1987)] by exchanging the order of the two expectations. The following result proved in the Appendix shows that the variance of $E[\hat{\sigma}^2|X_n]$ decays quite fast.

PROPOSITION 4.2. *If* (3.6) *holds and* $n_j/n \to \pi_j \in (0,1)$, $j = 1, \dots, p$, $\mathrm{Var}[E[\hat{\sigma}^2|X_n]] = O(c_n n^{-2})$ *with* c_n *defined in Theorem* 4.1.

For a complete discussion we also have to look at the variance of $\hat{\sigma}^2$. This depends on the fourth moments of (ε_t). The most important case is the one with Gaussian errors.

PROPOSITION 4.3. *If* (ε_t) *is Gaussian,* (3.6) *holds and* $n_j/n \to \pi_j \in (0,1)$, *then* $\mathrm{Var}[\hat{\sigma}^2|X_n] = O(c_n)$ *uniformly in* X_n, *where* c_n *has been defined in Theorem* 4.1.

PROOF. See the Appendix.

Hence when using the classical t- and F-tests for contrasts derived under independence the true level is presumably a little larger than we assume because the effective number of degrees of freedom is smaller. For a conclusive statement one would have to look also at the correlation between $\hat{\beta}_j(n) - \hat{\beta}_k(n)$ and $\hat{\sigma}^2$. In any case, asymptotically the level is correct.

5. Complete randomization. If (ε_t) is stationary, the model (3.2) is a special case of the general regression model with correlated errors. Moreover $n^{-1}\Sigma x_{t,j}x_{t+h,k}$ converges a.s. to $\pi_j\pi_k$ for $h \neq 0$ and to $\pi_j\delta_{jk}$ for $h = 0$, respectively. Hence the following theorem is an easy consequence of subsection 7.3 of Grenander and Rosenblatt (1957).

THEOREM 5.1. *If the spectrum* $f(\omega)$ *of* (ε_t) *is piecewise continuous and* $0 < c \leq f(\omega) \leq C$, *then* $nV_n(j,k|X_n)$ *converges a.s. to* $\sigma^2(\pi_j^{-1} + \pi_k^{-1})$ *and* $nV_n^*(j,k|X_n)$ *converges a.s. to* $\sigma^2((2\pi)^{-2}\int f(\omega)^{-1}\,d\omega)^{-1}(\pi_j^{-1} + \pi_k^{-1})$.

However if (ε_t) has long-range correlations, the spectrum is of the form (3.7) and thus the condition of Theorem 5.1 does not hold. Nevertheless we have the following:

THEOREM 5.2. *If* (3.7) *holds, then the conclusions of Theorem* 5.1 *continue to hold.*

The proof which uses a completely different approach and exploits the independence of the x_t's, is given in the Appendix.

Hence asymptotically $V_n(j,k|X_n)$ has the same behavior under complete and under restricted randomization. We conjecture that this is true also for $V_n^*(j,k|X_n)$.

It is well known [see, e.g., Rozanov (1967), Formula (10.40), page 104] that $\sigma^2((2\pi)^{-1}\int f(\omega)^{-1}\,d\omega)^{-1}$ is the variance of the best linear interpolation of ε_t based on ε_s, $s \neq t$. If the dependence is strong, this will be much smaller than $\mathrm{Var}[\varepsilon_t]$, that is, ordinary least squares can lose quite some efficiency. Some numerical values are given in Table 1 for the ARIMA $(0, H - 1/2, 0)$-model where the efficiency can be expressed with the gamma-function, see Theorem 1 of Hosking (1981). In practice the correlations of (ε_t) are unknown, so one

TABLE 1
Asymptotic efficiency of generalized versus ordinary least squares under complete randomization in the ARIMA(0, $H - 1/2$, 0)*-model*

H	0.6	0.7	0.8	0.9
efficiency	1.03	1.16	1.46	2.45

cannot expect to really achieve $V_n^*(j, k|X_n)$. Still, it should be possible to improve on the ordinary least squares estimator if n and H are large.

6. Randomized blocks with ordinary least squares.

6.1. *Asymptotic behavior of the variance.* First we give the exact value of the expectation and the order of magnitude of the variance of $V_n(j, k|X_n)$.

PROPOSITION 6.1. *Assume that* $\rho(t, s)$ *satisfies* (3.7). *Under* (3.5) *we have*:

(i) $E[V_n(j, k|X_n)] = (n_j^{-1} + n_k^{-1})\sigma^2(1 - 2(l - 1)^{-1}\Sigma_{t=1}^{l-1}\rho(t)(1 - t/l)) = (n_j^{-1} + n_k^{-1})\sigma_l^2$ *say.*

(ii) $\mathrm{Var}[nV_n(j, k|X_n)] = O(c_n)$, *where* c_n *has been defined in Theorem* 4.1.

PROOF. See the Appendix.

Proposition 6.1 implies the convergence of $nV_n(j, k|X_n)$ in probability. In fact we have even almost sure convergence.

THEOREM 6.1. *If* (3.7) *and* (3.5) *hold,* $nV_n(j, k|X_n)$ *converges almost surely to* $(\pi_j^{-1} + \pi_k^{-1})\sigma_l^2$ *where* $\pi_j = P[x_{t,j} = 1] = l_j/l$.

The proof is given in the Appendix. Hence with blockwise randomization the variance of a contrast depends on the correlations of the errors. If $\rho(t)$ is positive for all t, blocking decreases the variance. Moreover the decrease is larger for small blocks if the correlations decay monotonically.

LEMMA 6.1. *If* $\rho(t) > \rho(s) > 0$, $0 < t < s$, *then* σ_l^2 *is strictly increasing for* $l \geq 2$. *If* $\lim_{t\to\infty}\rho(t) = 0$, σ_l^2 *converges to* σ^2 *as* l *goes to infinity.*

PROOF. See Cochran (1946), pages 169–170.

Therefore blocking with large block size does not improve much over complete randomization. However, for small blocks the gain can be substantial as Table 2 shows. Comparing Tables 1 and 2 we see that ordinary least squares with small blocks is better than generalized least squares with complete randomization for the ARIMA(0, $H - 1/2$, 0)-model. This is however not true for any model with correlations decaying to zero monotonically. For instance,

TABLE 2
Asymptotic efficiency of blockwise versus complete randomization with ordinary least squares in the ARIMA(0, $H - 1/2$, 0)*-model. Block size* = l

H	0.6	0.7	0.8	0.9
eff for $l = 2$	1.12	1.33	1.75	3.00
eff for $l = 4$	1.09	1.25	1.59	2.61

TABLE 3
Asymptotic efficiency of generalized versus ordinary least squares under blockwise randomization in the ARIMA(0, $H - 1/2$, 0)*-model*

H	0.6	0.7	0.8	0.9
eff for $l = 2$	1.003	1.012	1.027	1.050
eff for $l = 4$	1.005	1.023	1.053	1.099

for a MA(1)-model with parameter α and $p = l = 2$, the efficiency of ordinary least squares with randomized blocks versus generalized least squares with complete randomization turns out to be $(1 - \alpha^2)/(1 + \alpha^2 - \alpha)$. For α going to one, this tends to zero.

One may ask at this point how much more efficiency can be gained by using generalized least squares (3.11) and/or different designs. With the same methods as used for Theorems 4.2 and 6.1 we can obtain the efficiency of generalized versus ordinary least squares for randomized blocks. For the ARIMA(0, $H - 1/2$, 0)-model the results are given in Table 3. It shows that ordinary least squares are sufficient for all practical purposes. If $p = 2$, $n_1 = n_2 = n/2$ and the spectral density $f(\omega)$ satisfies the conditions of Theorem 5.1 and takes its minimum at $\omega = \pi$, one can also show that the most efficient design alternates the two treatments. Moreover, ordinary and generalized least squares are equivalent for this design. However, this design is very sensitive to the presence of a monotone trend and thus should be avoided. We do not pursue the search for optimal designs any further.

6.2. *Asymptotic behavior of* $\hat{\sigma}^2_{\text{block}}$. Since we have exactly the same behavior as for $\hat{\sigma}^2$ in Section 4.1, we give only the main results. By generalizing Lemma 4.1 we obtain first the consistency of $\hat{\sigma}^2_{\text{block}}$.

THEOREM 6.2. *If* (ε_t) *is stationary and ergodic and* (3.7) *holds,* $\hat{\sigma}^2_{\text{block}} \to_P \sigma_l^2$ *both conditionally on the design and unconditionally.*

Hence also with randomized blocks the correlations of the errors can be neglected asymptotically when estimating the variance of a contrast because the inclusion of block effects makes the correct adjustment for the estimated

TABLE 4

Mean (upper figure), standard deviation (middle figure) and skewness (lower figure) of $\mathrm{Var}(\hat{\beta}_1(n) - \hat{\beta}_2(n)|X_n)/(n_1^{-1} + n_2^{-1})$ *for* $p = 2$, $n_1 = n_2 = n/2$. *The errors are discrete fractional noise with unit variance. Each value is estimated from a simulation of the design* X_n *under restricted randomization with* 400 *replicates*

	H	$n = 16$	$n = 64$
OLSE	0.7	0.857	0.934
		0.112	0.094
		0.85	0.11
	0.9	0.460	0.564
		0.142	0.125
		1.46	0.95
BLUE	0.7	0.789	0.806
		0.081	0.047
		1.37	0.54
	0.9	0.320	0.315
		0.066	0.030
		1.46	0.83

variance. Looking at finite n's we have by (3.13)

$$E[\hat{\sigma}^2_{\text{block}}|X_n] = \sigma^2 - \sigma^2(n-p-b+1)^{-1}$$

$$\times\left(2n/l\sum_{t=1}^{l-1}\rho(t)(1-t/l) + \sum_{t\neq s}\rho(t-s)\left(\sum_j n_j^{-1}x_{t,j}x_{s,j} - n^{-1}\right)\right). \tag{6.1}$$

Again by the validity of randomization $E[E[\hat{\sigma}^2_{\text{block}}|X_n]] = \sigma_l^2$. In the Appendix we prove the following:

PROPOSITION 6.2. $\mathrm{Var}[E[\hat{\sigma}^2_{block}|X_n]] = O(n^{-2}c_n)$.

7. Simulation results. We simulated 400 replicates of the design matrix X_n according to the restricted randomization (3.4) for sample sizes $n = 16$ and $n = 64$, number of treatments $p = 2$ and $p = 8$ and number of replicates $n_j \equiv n/p$. For the random errors ε_t we took zero mean, variance one and correlations $\rho(t,s) = (|t-s+1|^{2H} - 2|t-s|^{2H} + |t-s-1|^{2H})/2$, $H = 0.7$ and 0.9. In the Gaussian case this defines standardized fractional Gaussian noise. We computed the conditional variance of the ordinary least squares estimator (OLSE) and of the best linear unbiased estimator (BLUE) given the random design X_n according to (3.9) and (3.12). The behavior of these conditional variances is summarized in Tables 4 to 8.

In the case $p = 8$ there is a large number of possible contrasts to look at. Fortunately, by a simple symmetry argument the distributions of $V_n(j,k|X_n)$ or $V_n^*(j,k|X_n)$ are the same for any $j \neq k$. Hence we estimated the moments of $V_n(j,k|X_n)$ or $V_n^*(j,k|X_n)$ for a fixed $j \neq k$ by taking averages over all j and k. Besides these pairwise differences we looked also at orthogonal con-

TABLE 5

Mean, standard deviation and skewness of the variance of treatment differences given the design X_n for $p = 8$, $n_j = n/p$. Besides a fixed pair of treatments also the minimal and maximal variances among all pairs of treatments are given. Notation: $V_n^{st}(j,k) = \mathrm{Var}(\hat\beta_n(j) - \hat\beta_n(k)|X_n)/(n_j^{-1} + n_k^{-1})$, $V_n^{\max} = \max_{j \neq k} V_n^{st}(j,k)$ *and* $V_n^{\min} = \min_{j \neq k} V_n^{st}(j,k)$

		$V_n^{st}(j,k)$ for fixed $j \neq k$		$V_n^{\max}$		$V_n^{\min}$	
	H	$n = 16$	$n = 64$	$n = 16$	$n = 64$	$n = 16$	$n = 64$
OLSE	0.7	0.865	0.931	1.06	1.09	0.696	0.801
		0.100	0.079	0.072	0.069	0.030	0.027
		0.52	0.67	−0.22	0.70	0.38	−0.18
	0.9	0.455	0.572	0.702	0.844	0.276	0.393
		0.117	0.125	0.087	0.140	0.028	0.029
		0.79	1.11	−0.38	0.85	0.56	0.10
BLUE	0.7	0.819	0.811	0.971	0.899	0.676	0.729
		0.087	0.046	0.080	0.036	0.022	0.017
		0.57	0.43	0.39	0.43	0.41	−0.23
	0.9	0.359	0.326	0.498	0.389	0.243	0.271
		0.079	0.033	0.084	0.030	0.015	0.012
		0.89	0.68	0.29	0.76	0.39	0.07

TABLE 6

Mean, standard deviation and skewness of the variance and covariance of orthogonal contrasts given the design X_n with $p = 8$, $n_j = n/p$, $j = 1, \dots, p$

		$\mathrm{Var}(\hat\gamma_n(j)|X_n)p/n_j$ for j fixed		$\mathrm{Cov}(\hat\gamma_n(j), \hat\gamma_n(k)|X_n)p/(n_j n_k)^{1/2}$ for $j \neq k$ fixed	
	H	$n = 16$	$n = 64$	$n = 16$	$n = 64$
OLSE	0.7	0.863	0.933	0.000	0.001
		0.110	0.083	0.076	0.057
		0.80	0.72	−0.02	0.04
	0.9	0.453	0.574	0.001	0.001
		0.126	0.127	0.090	0.091
		1.05	1.02	0.02	0.00
BLUE	0.7	0.816	0.813	0.000	0.001
		0.093	0.049	0.064	0.034
		0.86	0.56	−0.04	0.06
	0.9	0.358	0.326	0.001	0.000
		0.082	0.033	0.056	0.023
		1.18	0.58	0.05	−0.04

TABLE 7
Mean, standard deviation and skewness of $\mathrm{Var}(\hat{\beta}_n(j) - \hat{\beta}_n(k)|X_n)/[(n_j^{-1} + n_k^{-1})E(\hat{\sigma}^2|X_n)]$ *for* $j \neq k$ *fixed*, $n_j = n/p$, $j = 1, \dots, p$. *The estimator used is* OLSE

	$p = 2$		$p = 8$	
	$n = 16$	$n = 64$	$n = 16$	$n = 64$
$H = 0.7$	1.01	1.01	1.00	1.00
	0.143	0.098	0.067	0.034
	0.97	0.89	0.51	0.28
$H = 0.9$	1.00	1.02	1.01	1.00
	0.313	0.231	0.150	0.091
	1.45	1.15	0.58	0.57

trasts $\gamma_j = \sum_{i=1}^{8} \lambda_{ji} \beta_i$, $j = 1, \dots, 7$, where the vectors $(\lambda_{j1}, \dots, \lambda_{j8})^T$ are orthogonal and have elements ± 1. This corresponds to main effects and interactions in a 2^3 factorial. Again the distribution of $\mathrm{Var}(\hat{\gamma}_j(n)|X_n)$ and $\mathrm{Cov}(\hat{\gamma}_j(n), \hat{\gamma}_k(n)|X_n)$ are the same for all j and $j \neq k$, respectively.

The asymptotic values of $nV_n^*(j, k|X_n)/(1/n_j + 1/n_k)$, given in Theorem 5.1, are equal to 0.810 for $H = 0.7$ and 0.317 for $H = 0.9$. Tables 4, 5 and 6 illustrate that already for short series the expected value of the variance for the BLUE is very close to the asymptotic limit. Also the standard deviation of $nV_n^*(j, k|X_n)$ decreases fast, probably with the rate $n^{-1/2}$. On the other hand, $E(nV_n(j, k|X_n))/(1/n_j + 1/n_k)$ is quite far from the asymptotic value of 1, as predicted by Proposition 4.1(i). In particular for $H = 0.9$ the bias is large. Also the standard deviation decreases very slowly. Note furthermore that the skewness of the BLUE is decreasing with increasing sample size. Less clear is the behavior of the simulated skewness for the variance of the OLSE. The s.d. of the empirical skewness of 400 Gaussian random variables is 0.123, that is, most skewness values in Tables 4–8 are significant. It is an open problem if $V_n(j, k|X_n)$, normalized by a factor of the order $c_n^{1/2}$ given by Theorem 4.1, is asymptotically normal or if it has another limit distribution. In principle,

TABLE 8
Mean, standard deviation and skewness of $\mathrm{Var}(n^{-1}\sum_{t=1}^{n} y_t)/n^{-1}E(\hat{\sigma}^2|X_n)$ *for* $n_j = n/p$, $j = 1, \dots, p$. *The estimator used is* OLSE

	$p = 2$		$p = 8$	
	$n = 16$	$n = 64$	$n = 16$	$n = 64$
$H = 0.7$	3.51	5.66	3.51	5.66
	0.033	0.009	0.109	0.021
	0.99	0.94	0.51	0.30
$H = 0.9$	20.2	48.6	20.4	48.5
	0.423	0.178	1.42	0.494
	1.45	1.19	0.59	0.58

$V_n(j,k|X_n)$ belongs to the class of randomly associated arrays considered by Barbour and Eagleson (1986), but we were unable to verify any of their conditions for asymptotic distributions.

The smaller variability of the conditional variance for the BLUE can also be seen by comparing $\Delta_n^* = \max_{j \neq k} V_n^*(j,k|X_n) - \min_{j \neq k} V_n^*(j,k|X_n)$ with $\Delta_n = \max_{j \neq k} V_n(j,k|X_n) - \min_{j \neq k} V_n(j,k|X_n)$ (Table 5). For the BLUE, increasing the sample size from $n = 16$ to $n = 64$ leads to a decrease of the mean of this difference by an approximate factor of 1/2. This suggests a decay of the order $n^{-1/2}$. On the other hand, for the OLSE and $H = 0.7$ the decay is much slower. For $H = 0.9$ the mean of Δ_n is even approximately the same for $n = 16$ and $n = 64$. This means that for relatively small samples the conditional variance of the OLSE of a contrast varies considerably, depending on which design matrix we obtained from the randomization. For the BLUE the variance is less sensitive to the (randomized) choice of the treatment assignments. In order to obtain design matrices for which the conditional LS-variance $V_n(j,k)$ is more stable when n is not very large, one might use constrained randomization; see Bailey (1985).

Table 7 illustrates how the biases of $V_n(j,k|X_n)$ and of $E[\hat{\sigma}^2|X_n]$ compensate each other, compare Propositions 4.1 and 4.2. The mean of the ratio of the two quantities is practically equal to one even for $n = 16$. The value of H or the number of treatments does not seem to play any role here. Table 8 shows that the situation is completely different for the arithmetic mean. There the increase in the variance due to the dependence of the errors and the bias of $E[\hat{\sigma}|X_n]$ reinforce each other. Hence the ratio of the two quantities is huge even for $n = 16$ and $H = 0.7$. This emphasizes once more the fundamental difference between contrasts and absolute constants.

APPENDIX

Proofs. Without loss of generality we assume $\sigma^2 = 1$ throughout.

Proof of Proposition 4.1. We put $\xi_t = x_{t,j}/n_j - x_{t,k}/n_k$. Because $\sum \xi_t^2 = (n_j^{-1} + n_k^{-1})$, it follows from (3.9) that $V_n(j,k|X_n) = (n_j^{-1} + n_k^{-1}) + \sum_{t \neq s} \rho(t,s)\xi_t \xi_s$. The randomization (3.4) implies that for $t \neq s$,

$$E[\xi_t \xi_s] = n^{-1}(n-1)^{-1}\left\{n_j^{-2} n_j(n_j - 1) + n_k^{-2} n_k(n_k - 1) - 2n_j^{-1} n_k^{-1} n_j n_k\right\}$$
$$= -n^{-1}(n-1)^{-1}\left(n_j^{-1} + n_k^{-1}\right).$$

This proves (i). For (ii) we observe that

$$\begin{aligned}\operatorname{Var}[V_n(j,k|X_n)] &= 2\sum{}'\rho(t,s)^2 \operatorname{Var}[\xi_t \xi_s] \\ &\quad + 4\sum{}'\rho(t,s)\rho(s,u)\operatorname{Cov}[\xi_t \xi_s, \xi_s \xi_u] \\ &\quad + \sum{}'\rho(t,s)\rho(u,v)\operatorname{Cov}[\xi_t \xi_s, \xi_u \xi_v].\end{aligned}$$

Now a lengthy but straightforward calculation shows that for t, s, u, v all different

$$\operatorname{Var}[\xi_t\xi_s] = n^{-2}\left(n_j^{-1} + n_k^{-1}\right)^2 + O(n^{-5}),$$

$$\operatorname{Cov}[\xi_t\xi_s, \xi_s\xi_u] = -n^{-3}\left(n_j^{-1} + n_k^{-1}\right)^2 + O(n^{-6}),$$

$$\operatorname{Cov}[\xi_t\xi_s, \xi_u\xi_v] = 2n^{-4}\left(n_j^{-1} + n_k^{-1}\right)^2 + O(n^{-7}).$$

Together this proves (ii). □

PROOF OF THEOREM 4.1. By (3.6) $|\Sigma'\rho(t, s)| \le \text{const.}\, n\sum_{t=1}^{n-1} t^{2H-2} \le \text{const.}\, n^{2H}$, and

$$\begin{aligned}\sum{}' \rho(t,s)^2 &\le \text{const.}\, n \sum_{t=1}^{n-1} t^{4H-4} \le \text{const.}\, n \quad \text{if } H < \tfrac{3}{4},\\ &\le \text{const.}\, n \log n \quad \text{if } H = \tfrac{3}{4},\\ &\le \text{const.}\, n^{4H-2} \quad \text{if } H > \tfrac{3}{4}.\end{aligned}$$

Similarly

$$\begin{aligned}\left|\sum{}' \rho(t,s)\rho(s,u)\right| &\le \text{const.} \sum{}' |\rho(t,s)| \sum_{u=1}^{n-1} u^{2H-2}\\ &\le \text{const.}\, n \left(\sum_{u=1}^{n-1} u^{2H-2}\right)^2 \le \text{const.}\, n^{4H-1}.\end{aligned}$$

Finally $|\Sigma'\rho(t, s)\rho(u, v)| \le \text{const.}(n\sum_{t=1}^{n-1} t^{2H-2})^2 \le \text{const.}\, n^{4H}$. Hence Theorem 4.1 follows from Proposition 4.1. □

PROOF OF LEMMA 4.1. Because $x_{t,j} \in \{0, 1\}$,

$$\begin{aligned}E\left[(n-p)^{-1}\sum_j n_j^{-1}\left(\sum_t \varepsilon_t x_{t,j}\right)^2 |X_n\right] &\le (n-p)^{-1}\sum_j n_j^{-1}\sum_{t,s}|\rho(t,s)|\\ &\le \text{const.}\, n^{2H-1}\sum_j n_j^{-1} = O(n^{2H-2}). \quad \square\end{aligned}$$

PROOF OF PROPOSITION 4.2. We put $\eta_{t,s} = (n-p)^{-1}\Sigma n_j^{-1} x_{t,j} x_{s,j}$. Then $E[\eta_{t,s}] = (n-p)^{-1} n^{-1}(n-1)^{-1}\Sigma n_j^{-1} n_j(n_j - 1) = n^{-1}(n-1)^{-1}$, $t \ne s$, and by a lengthy but straightforward calculation we find for t, s, u, v all different

$$\operatorname{Var}[\eta_{t,s}] = n^{-4}(p-1) + O(n^{-5}),$$

$$\operatorname{Cov}[\eta_{t,s}, \eta_{s,u}] = -n^{-5}(p-1) + O(n^{-6}),$$

$$\operatorname{Cov}[\eta_{t,s}, \eta_{u,v}] = 2(p-1)n^{-6} + O(n^{-7}).$$

The proof is then completed with the arguments used above for Proposition 4.1 and Theorem 4.1. □

PROOF OF PROPOSITION 4.3. By the well-known formula for the fourth moment of a Gaussian random variable we have

$$\operatorname{Var}\left[\sum_t \varepsilon_t^2\right] = 2\sum_{t,s}\rho(t,s)^2,$$

$$\operatorname{Var}\left[\left(\sum_t \varepsilon_t x_{t,j}\right)^2 | X_n\right] = 2\operatorname{Var}\left[\sum_t \varepsilon_t x_{t,j} | X_n\right]^2 = 2\left(\sum_{t,s} x_{t,j}x_{s,j}\rho(t,s)\right)^2$$

$$\leq 2\left(\sum |\rho(t,s)|\right)^2,$$

$$\left|\operatorname{Cov}\left[\sum_t \varepsilon_t^2, \left(\sum_t \varepsilon_t x_{t,j}\right)^2 | X_n\right]\right| = \left|2\sum_{t,s,u} x_{t,j}x_{s,j}\rho(t,u)\rho(s,u)\right|$$

$$\leq 2\sum_{t,s,u} |\rho(t,u)||\rho(s,u)|.$$

Now we apply again the arguments of the proof of Theorem 4.1. □

The proof of Theorem 5.2 is more difficult than the previous ones. We break it up into several parts.

LEMMA A.1. $\lim nV_n(j,k|X_n) = (\pi_j^{-1} + \pi_k^{-1})$.

PROOF. As above we introduce

$$\xi_t = x_{t,j}/n_j - x_{t,k}/n_k.$$

Because n_j is random, it is much easier if we can replace ξ_t by

$$\theta_t = n^{-1}(x_{t,j}/\pi_j - x_{t,k}/\pi_k).$$

Because the θ_t's are independent with expectation zero and variance $n^{-2}(\pi_j^{-1} + \pi_k^{-1})$,

$$M_n = n^2\sum_{t,s=1}^{n} \theta_t\rho(t-s)\theta_s - n\left(\pi_j^{-1} + \pi_k^{-1}\right)$$

is a martingale with respect to $\mathscr{F}_n = \sigma(\theta_1,\dots,\theta_n)$. Moreover

$$E\left[(M_n - M_{n-1})^2\right] = \operatorname{Var}\left[n^2\theta_n^2\right] + 4E\left[n^2\theta_n^2\right]^2\sum_{t=1}^{n-1}\rho(n-t)^2 \leq \text{const.}\, nc_n.$$

Therefore by monotone integration,

$$E\left[\sum n^{-2}E\left[(M_n - M_{n-1})^2 | \mathscr{F}_{n-1}\right]\right] = \sum n^{-2}E\left[(M_n - M_{n-1})^2\right] < \infty.$$

The strong law of large numbers for martingales [see, e.g., Stout (1974), Theorem 3.3.1] implies therefore that $n^{-1}M_n = n\sum_{t,s=1}^{n}\theta_t\rho(t-s)\theta_s - (\pi_j^{-1} + \pi_k^{-1})$ converges to zero a.s.

Finally we have to show that

$$V_n(j,k|X_n) - \sum_{t,s=1}^{n} \theta_t \rho(t-s)\theta_s = \xi^T \Sigma_n \xi - \theta^T \Sigma_n \theta = o(n^{-1}) \quad \text{a.s.}$$

Observe that

$$\begin{aligned} |\xi^T\Sigma_n\xi - \theta^T\Sigma_n\theta| &\le \left((\xi-\theta)^T\Sigma_n(\xi-\theta)\right)^{1/2}\left((\xi^T\Sigma_n\xi)^{1/2} + (\theta^T\Sigma_n\theta)^{1/2}\right) \\ &\le 2\left((\xi-\theta)^T\Sigma_n(\xi-\theta)\right)^{1/2}\left(\theta^T\Sigma_n\theta\right)^{1/2} \\ &\quad + (\xi-\theta)^T\Sigma_n(\xi-\theta). \end{aligned}$$

By the argument above $\theta^T\Sigma_n\theta = O(n^{-1})$ a.s. By (3.6) the largest eigenvalue of Σ_n is bounded by const. n^{2H-1}. Hence

$$\begin{aligned} (\xi-\theta)^T\Sigma_n(\xi-\theta) &\le \text{const.}\, n^{2H-1}(\xi-\theta)^T(\xi-\theta) \\ &\le \text{const.}\, n^{2H}\left(\left(n_j^{-1} - (n\pi_j)^{-1}\right)^2 + \left(n_k^{-1} - (n\pi_k)^{-1}\right)^2\right), \end{aligned}$$

which is $O(n^{2H-3}(\log\log(n))^2) = o(n^{-1})$ by the law of the iterated logarithm. This completes the proof of Lemma A.1. □

Lemma A.2. $\limsup nV_n^*(j,k|X_n) \ge ((2\pi)^{-2}\int f(\omega)^{-1}\,d\omega)^{-1}(\pi_j^{-1} + \pi_k^{-1})$.

Proof. Let (g_m) be the sequence of spectral densities $g_m(\omega) = \min(f(\omega), m)$. By $\Sigma_{m,n}$ we denote the corresponding covariance matrix of n consecutive observations. Now $g_m \le f$ implies $\Sigma_{m,n} \le \Sigma_n$, the last inequality being in the positive definite sense. By standard arguments it follows that $(X_n^T\Sigma_{m,n}^{-1}X_n)^{-1} \le (X_n^T\Sigma_n^{-1}X_n)^{-1}$ whence $\liminf nV_n^*(j,k|X_n) \ge \lim n\lambda^T(X_n\Sigma_{m,n}^{-1}X_n)^{-1}\lambda$. This limit exists a.s. by Theorem 5.1 and equals $((2\pi)^{-2}\int g_m(\omega)^{-1}\,d\omega)^{-1}(\pi_j^{-1} + \pi_k^{-1})$. Letting m tend to infinity completes the proof. □

Lemma A.3. $\limsup nV_n^*(j,k|X_n) \le ((2\pi)^{-2}\int f(\omega)^{-1}\,d\omega)^{-1}(\pi_j^{-1} + \pi_k^{-1})$.

Proof. This is the hard part because we cannot approximate f from above. We are going to approximate the best linear unbiased estimator $\tilde\beta(n)$ by $\tilde\beta^W(n) = (X_n^TW_nX_n)^{-1}(X_n^TW_ny)$, where $W_n = \{w(t-s);\ 1 \le t,s \le n\}$ with $w(t) = w(-t)$, $w(t) = 0$ $(|t| > q)$, $\hat w(\omega) = (2\pi)^{-1}\Sigma_t w(t)e^{it\omega} > 0$. This way we avoid approximating Σ_n^{-1} which is very delicate, compare Bleher (1981).

Because $\tilde\beta(n)$ is optimal, we have

$$\begin{aligned} nV_n^*(j,k|X_n) &\le n\,\mathrm{Var}\left(\tilde\beta_j^W(n) - \tilde\beta_k^W(n)|X_n\right) \\ &= n\lambda^T\left(X_n^TW_nX_n\right)^{-1}X_n^TW_n\Sigma_nW_nX_n\left(X_n^TW_nX_n\right)^{-1}\lambda. \end{aligned}$$

Using Lemma A.4 and approximating $f(\omega)^{-1}$ by a sequence of positive trigonometric polynomials $\hat w(\omega)$ completes the proof. □

Lemmas A.1–A.3 imply Theorem 5.2. Finally we show the following:

LEMMA A.4. *$n\lambda^T(X_n^T W_n X_n)^{-1} X_n^T W_n \Sigma_n W_n X_n (X_n^T W_n X_n)^{-1}\lambda$ converges a.s. to $w(0)^{-2}(2\pi)^2 \int \hat{w}(\omega)^2 f(\omega)\, d\omega(\pi_j^{-1} + \pi_k^{-1})$. ($W_n$, w and $\hat{w}$ as defined above.)*

PROOF. This is basically the same proof as the convergence of $nV_n(j, k|X_n)$. Let γ_t be defined by

$$\gamma_t = \left(X_n\left(X_n^T W_n X_n\right)^{-1}\lambda\right)_t w(0)$$

and let θ_t be as above. Again $M_n = n^2\theta^T W_n \Sigma_n W_n \theta - \operatorname{tr}(W_n \Sigma_n W_n)(\pi_j^{-1} + \pi_k^{-1})$ is a martingale. Because

$$|(W_n \Sigma_n W_n)_{t,s}| \le \left(\sum_u |w(u)|\right)^2 \max_{-2q \le j \le 2q} |\rho(t - s + j)| \le \text{const.}|t - s|^{2H-2},$$

$n^{-1}M_n$ converges a.s. to zero. Moreover for $q < t < n - q$

$$(W_n \Sigma_n W_n)_{t,t} = \sum_{s,u} w(s) w(u) \rho(s - u) = (2\pi)^2 \int \hat{w}(\omega)^2 f(\omega)\, d\omega.$$

Hence $n^{-1} \operatorname{tr}(W_n \Sigma_n W_n)$ converges to $(2\pi)^2 \int \hat{w}(\omega)^2 f(\omega)\, d\omega$.

Finally we note that the largest eigenvalue of $W_n \Sigma_n W_n$ is of the same order as the one of Σ_n. Hence we only need to show that $\sum_t (\gamma_t - \theta_t)^2 = o(n^{-2H})$ a.s. But

$$\begin{aligned}\left(X_n^T W_n X_n\right)_{jk} &= w(0) \sum_{t=1}^{n} x_{t,j} x_{t,k} + \sum_{s=1}^{q} w(s) \sum_{t=1}^{n-s} (x_{t,j} x_{t+s,k} + x_{t,k} x_{t+s,j}) \\ &= n\left\{w(0)\delta_{jk}\pi_j + 2\sum w(s)\pi_j\pi_k\right\} + O\left(n^{1/2} \log\log(n)\right) \\ &= nA_{jk} + O\left(n^{1/2}\log\log(n)\right) \quad \text{say.}\end{aligned}$$

It is straightforward to check that $(A^{-1}\lambda)_i = w(0)^{-1}(\delta_{ij}\pi_j^{-1} - \delta_{ik}\pi_k^{-1})$. Together this implies $|\gamma_t - \theta_t| = O(n^{-3/2} \log\log(n))$ a.s. (uniformly in t) which completes the proof of Lemma A.4. □

PROOF OF PROPOSITION 6.1 AND THEOREM 6.1. Again let $\xi_t = x_{t,j}/n_j - x_{t,k}/n_k = b^{-1}(x_{t,j}/l_j - x_{t,k}/l_k)$ and

$$Z(i_1, i_2) = \sum_{t \in B_{i_1}} \sum_{s \in B_{i_2}} \rho(t - s)\xi_t\xi_s, \qquad 1 \le i_1, i_2 \le b.$$

By the independence of different blocks $E[Z(i_1, i_2)] = 0$ for $i_1 \ne i_2$. If $t \ne s$ are in the same block

$$E[\xi_t\xi_s] = -n^{-1}(l - 1)^{-1}\left(n_j^{-1} + n_k^{-1}\right).$$

Hence $E[Z(i,i)] = (n_j^{-1} + n_k^{-1})b^{-1}(1 - 2(l - 1)^{-1}\sum_{t=1}^{l-1}\rho(t)(1 - t/l))$. Because $V_n(j, k|X_n) = \sum_{i_1, i_2} Z(i_1, i_2)$, this proves (i) of Proposition 6.1.

Again by the independence of different blocks $\mathrm{Var}[V_n(j, k|X_n)] = \Sigma_{i_1, i_2} \mathrm{Var}(Z(i_1, i_2))$. Because $|\xi_t| \le \text{const.}\, b^{-1}$, we have by (3.6)

$$\mathrm{Var}\big(Z(i_1, i_2)\big) \le \text{const.}\, b^{-4}|i_1 - i_2|^{4H-4}.$$

From this (ii) of Proposition 6.1 follows easily. Finally, for Theorem 6.1 we note that

$$M_b = b^2 \sum_{i_1=1}^{b} \sum_{i_2=1}^{b} Z(i_1, i_2) - l^{-1}\sigma_l^2\big(\pi_j^{-1} + \pi_k^{-1}\big)$$

is a martingale with respect to $\mathscr{F}_b = \sigma(x_1, \dots, x_{lb})$. The condition for the strong law of large numbers is easily checked. Details are left to the reader. □

PROOF OF PROPOSITION 6.2. Denote $\eta_{t,s} = \Sigma_j n_j^{-1} x_{t,j} x_{s,j} - n^{-1}$, $t \ne s$, $\eta_{t,t} = 0$, and introduce

$$Z(i_1, i_2) = \sum_{t \in B_{i_1}} \sum_{s \in B_{i_2}} \rho(t - s)\eta_{t,s}.$$

For t and s in different blocks, we have $E[\eta_{t,s}|x_t] = \Sigma_j x_{t,j}/n - n^{-1} = 0$. Similarly $E[Z(i_1, i_2)|x_t, t \notin B_{i_2}] = 0$ for $i_1 \ne i_2$. Hence by (6.1)

$$\begin{aligned}\mathrm{Var}\big[E\big[\hat\sigma^2_{\text{block}}|X_n\big]\big] &\sim n^{-2}\,\mathrm{Var}\left(\sum_{i_1=1}^{b}\sum_{i_2=1}^{b} Z(i_1, i_2)\right)\\ &= n^{-2} \sum_{i_1=1}^{b}\sum_{i_2=1}^{b} \mathrm{Var}\big(Z(i_1, i_2)\big).\end{aligned}$$

Because $|\eta_{t,s}| \le \text{const.}\, n^{-1}$ we can complete the proof with the same argument as for Proposition 6.1. □

Acknowledgment. We wish to thank two referees for their careful and critical reading of the first version. Their comments led to an improvement in the presentation of our results.

REFERENCES

BAILEY, R. A. (1985). Restricted randomization versus blocking. *Internat. Statist. Rev.* **53** 171–182.

BAILEY, R. A. and ROWLEY, C. A. (1987). Valid randomization. *Proc. Roy. Soc. London Ser. A* **410** 105–124.

BARBOUR, A. D. and EAGLESON, G. K. (1986). Random association of symmetric arrays. *Stochastic Anal. Appl.* **4** 239–281.

BERAN, J. (1986). Estimation, testing and prediction for self-similar and related processes. Ph.D. dissertation, ETH, Zürich.

BERAN, J. (1989). A test of location for data with slowly decaying serial correlations. *Biometrika* **76** 261–269.

BERAN, J. (1992). Statistical methods for data with long-range dependence (with discussion). *Statist. Sci.* **7** 404–427.

BERENBLUT, I. I. and WEBB, G. J. (1974). Experimental designs in the presence of autocorrelated errors. *Biometrika* **61** 427–437.

BLEHER, P. M. (1981). Inversion of Toeplitz matrices. *Trans. Moscow Math. Soc.* **40** 201–229.

CHENG, C.-S. (1983). Construction of optimal balanced incomplete block designs for correlated observations. *Ann. Statist.* **11** 240–246.

COCHRAN, W. G. (1946). Relative accuracy of systematic and stratified random samples for a certain class of populations. *Ann. Math. Statist.* **17** 164–177.

FISHER, R. A. (1925). *Statistical Methods for Research Workers*. Oliver and Boyd, Edinburgh.

GRAF, H. P. (1983). Long-range correlations and estimation of the self-similarity parameter. Ph.D. dissertation, ETH, Zürich.

GRAF, H. P., HAMPEL, F. R. and TACIER, J. (1984). The problem of unsuspected serial correlations. *Robust and Nonlinear Time Series Analysis. Lecture Notes in Statist.* **26** 127–145. Springer, New York.

GRENANDER, U. and ROSENBLATT, M. (1957). *Analysis of Stationary Time Series*. Wiley, New York.

HAMPEL, F. R. (1987). Data analysis and self-similar processes, with discussion. In *Proceedings of the 46th Session of the ISI* **52** (Book 4) 235–264.

HAMPEL, F. R., RONCHETTI, E. M., ROUSSEEUW, P. J. and STAHEL, W. A. (1986). *Robust Statistics: The Approach Based on Influence Functions*. Wiley, New York.

HOSKING, J. R. M. (1981). Fractional differencing. *Biometrika* **68** 165–176.

JEFFREYS, H. (1939). *Theory of Probability*. Clarendon Press, Oxford. (2nd ed. published 1948; 3rd ed. published 1961).

JENKINS, G. M. and CHANMUGAN, J. (1962). The estimation of slope when the errors are autocorrelated. *J. Roy. Statist. Soc. Ser. B* **24** 199–214.

KIEFER, J. and WYNN, H. P. (1981). Optimum balanced block and latin square designs for correlated observations. *Ann. Statist.* **9** 737–757.

KIEFER, J. and WYNN, H. P. (1984). Optimum and minimax exact treatment designs for one-dimensional autoregressive error processes. *Ann. Statist.* **12** 414–450.

KOLMOGOROV, A. N. (1940). Wienersche Spiralen und einige andere interessante Kurven im Hilbertschen Raum. *C.R. Acad. Sci. URSS (N.S.)* **26** 115–118.

KÜNSCH, H. R. (1987). Statistical aspects of self-similar processes. In *Proceedings of the First World Congress of the Bernoulli Society* (Yu. Prohorov and V. V. Sazonov, eds.) **1** 67–74. VNU Science Press, Utrecht.

MANDELBROT, B. B. and VAN NESS, J. W. (1968). Fractional Brownian motions, fractional noises and applications. *SIAM Rev.* **10** 422–437.

MANDELBROT, B. B. and WALLIS, J. R. (1968). Noah, Joseph, and operational hydrology. *Water Resources Research* **4** 909–918.

MANDELBROT, B. B. and WALLIS, J. R. (1969). Some long-run properties of geophysical records. *Water Resources Research* **5** 321–340.

MARTIN, R. J. (1986). On the design of experiments under spatial correlation. *Biometrika* **73** 247–277.

MORGAN, J. P. and CHAKRAVARTI, I. M. (1988). Block designs for first and second order neighbour correlations. *Ann. Statist.* **16** 1206–1224.

NEWCOMB, S. (1895). Astronomical Constants (The Elements of the Four Inner Planets and the Fundamental Constants of Astronomy). Supplement to the *American Ephemeris and Nautical Almanac for 1897*. U.S. Government Printing Office, Washington, D.C.

PEARSON, K. (1902). On the mathematical theory of errors of judgement, with special reference to the personal equation. *Philos. Trans. Roy. Soc. London Ser. A* **198** 235–299.

PERCIVAL, D. B. (1985). On the sample mean and variance of a long memory process. Technical Report 69, Dept. Statistics, Univ. Washington, Seattle.

ROZANOV, YU.A. (1967). *Stationary Random Processes*. Holden-Day, San Francisco.

SINAI, YA.G. (1976). Self-similar probability distributions. *Theory Probab. Appl.* **21** 64–80.

STOUT, W. F. (1974). *Almost Sure Convergence*. Academic, New York.

STUDENT (1927). Errors of routine analysis. *Biometrika* **19** 151–164.

VOSS, R. V. (1979). $1/f$ (Flicker) noise: a brief review. In *Proceedings of the 33rd Annual Frequency Control Symposium* 40–46. IEEE, New York.

WILLIAMS, R. M. (1952). Experimental design for serially correlated observations. *Biometrika* **39** 151–167.

YAJIMA, Y. (1988). On estimation of a regression model with long-memory stationary errors. *Ann. Statist.* **16** 791–807.

YAJIMA, Y. (1991). Asymptotic properties of the LSE in a regression model with long-memory stationary errors. *Ann. Statist.* **19** 158–177.

ZYGMUND, A. (1959). *Trigonometric Series*. Cambridge University Press.

J. BERAN
DEPARTMENT OF BIOSTATISTICS
ISPM, UNIVERSITY OF ZÜRICH
SUMATRASTR. 30, CH-8006 ZÜRICH
SWITZERLAND

F. HAMPEL
H. KÜNSCH
SEMINAR FÜR STATISTIK
ETH ZENTRUM
CH-8092 ZÜRICH
SWITZERLAND