Biometrika (2025), **112**, 2, asaf025
https://doi.org/10.1093/biomet/asaf025
Advance Access publication 7 April 2025

On the minimax robustness against correlation and heteroscedasticity of ordinary least squares among generalized least squares estimates of regression

By DOUGLAS P. WIENS

Department of Mathematical and Statistical Sciences, University of Alberta, Edmonton, Alberta T6G 2G1, Canada
doug.wiens@ualberta.ca

Summary

We revisit a result according to which certain functions of covariance matrices are maximized at scalar multiples of the identity matrix. In a statistical context in which such functions measure loss, this says that the least favourable form of dependence is in fact independence, so that a procedure optimal for independent and identically distributed data can be minimax. In particular, the ordinary least squares estimate of a correctly specified regression response is minimax among generalized least squares estimates, when the maximum is taken over certain classes of error covariance structures and the loss function possesses a natural monotonicity property. In regression models whose response function is possibly misspecified, ordinary least squares is minimax if the design is uniform on its support, but this often fails otherwise. An investigation of the interplay between minimax generalized least squares procedures and minimax designs leads us to extend, to robustness against dependencies, an existing observation: that robustness against model misspecifications is increased by splitting replicates into clusters of observations at nearby locations.

Some key words: Design; Induced matrix norm; Loewner ordering; Particle swarm optimization; Robustness.

1. Introduction

When carrying out a study, whether observational or designed, calling for a regression analysis, the investigator may be faced with questions regarding possible correlations or heteroscedasticity within the data. If there are such departures from the assumptions underlying the use of the ordinary least squares (OLS) estimates of the regression parameters, then the use of generalized least squares (GLS) might be called for. In its pure form, as envisioned by Aitken (1935), this calls for the use of the inverse of the covariance matrix C, i.e., the precision matrix, of the random errors. This is inconvenient, since C is rarely known. If a consistent estimate $\hat{C}^{-1}$ of the precision matrix exists then one can employ 'feasible generalized least squares' estimation; see, e.g., Fomby et al. (1984). An example is the Cochrane–Orcutt procedure (Cochrane & Orcutt, 1949), which can be applied iteratively in first-order autoregressive AR(1) models. Otherwise, a positive definite 'pseudo-precision' matrix P might be employed. With data y and design matrix X, this leads to the estimate

$$\hat{\theta}_{\text{gls}} = \arg\min_\theta \left|\left|P^{1/2}(y - X\theta)\right|\right|^2 = (X^{\mathrm{T}}PX)^{-1}X^{\mathrm{T}}Py. \tag{1}$$

Wiens (2024b) used a lemma, restated below as Lemma 1, to show that certain commonly employed loss functions, taking covariance matrices as their arguments and increasing with respect to the

Loewner ordering by positive semidefiniteness, are maximized at scalar multiples of the identity matrix. This has the somewhat surprising statistical interpretation that the least favourable form of dependence is in fact independence. The lemma was used to show that the assumption of uncorrelated and homoscedastic errors at the design stage of an experiment is in fact a minimax strategy, within broad classes of alternate covariance structures.

In this article we study the implications of the lemma in the problem of choosing between OLS and GLS estimation methods. We first show that, when the form of the regression response is accurately modelled, then it can be safe, and indeed optimal (in a minimax sense), to ignore possible departures from independence and homoscedasticity, varying over certain large classes of such departures. This is because the common functions measuring the loss incurred by GLS, when the covariance matrix of the errors is C, are maximized when C is a multiple of the identity matrix. But in that case the best GLS estimate is OLS, i.e., OLS is a minimax procedure.

We then consider the case of misspecified response models, in which bias becomes a component of the integrated mean squared prediction error (IMSPE). The IMSPE is maximized over C and over the departures from the fitted linear response model. We show that, if a GLS with pseudo-precision matrix P is employed then the variance component of this maximum continues to be minimized by $P = I$, i.e., by OLS, but the bias generally does not and, depending upon the design, OLS can fail to be a minimax procedure. We show however that if the design is uniform on its support, i.e., places an equal number of observations at each of a number of points, then OLS is minimax. Otherwise, OLS can fail to be minimax when the design emphasizes bias reduction over variance reduction to a sufficiently large extent.

We also construct minimax designs, minimizing the maximum IMSPE over the design, and combine them with minimax choices of P. These designs are often uniform on their supports, and so OLS is a minimax procedure in this context. The design uniformity is typically attained by replacing the replicates that are a feature of 'classically optimal' designs minimizing variance alone by clusters of observations at nearby design points.

The contribution of Wiens (2024b) was to establish that designs, tailored for use with OLS estimation, have minimax properties within the class of designs. The current work establishes that, regardless of the design, OLS estimates are minimax within the class of GLS estimates, when the response is correctly specified. As well, for misspecified models, a summary of our findings is that, if a sensible design is chosen then OLS is at least 'almost' a minimax GLS procedure, often exactly so. We conclude that, for Loewner-increasing loss functions, and for covariance matrices C varying over the classes covered by Lemma 1, the simplicity of OLS makes it a robust and attractive alternative to GLS.

2. A USEFUL LEMMA

Suppose that $\|\cdot\|_M$ is a matrix norm, induced by a vector norm $\|\cdot\|_V$, i.e., $\|C\|_M = \sup_{\|x\|_V=1} \|Cx\|_V$. Special cases are the spectral radius $\|C\|_E$ (this is the maximum eigenvalue if C is a covariance matrix) and the maximum absolute row sum $\|C\|_1 = \max_i \sum_j |c_{ij}|$. Now suppose that the loss function in a statistical problem is $\mathcal{L}(C)$, where C is an $n \times n$ covariance matrix and $\mathcal{L}(\cdot)$ is nondecreasing in the Loewner ordering: $A \preceq B \Rightarrow \mathcal{L}(A) \leqslant \mathcal{L}(B)$. Here $A \preceq B$ means that $B - A \succeq 0$, i.e., is positive semidefinite.

The following lemma is established by Wiens (2024b).

LEMMA 1. *For $\eta^2 > 0$, covariance matrix C and induced norm $\|C\|_M$, define*

$$\mathcal{C}_M = \{C \mid C \succeq 0 \text{ and } \|C\|_M \leqslant \eta^2\}.$$

For the norm $\|\cdot\|_E$, an equivalent definition is

$$\mathcal{C}_E = \{C \mid 0 \preceq C \preceq \eta^2 I_N\}.$$

Then

(i) *in any such class* $\mathcal{C}_M$, $\max_{\mathcal{C}_M} \mathcal{L}(C) = \mathcal{L}(\eta^2 I_n)$, *and*
(ii) *if* $\mathcal{C}' \subseteq \mathcal{C}_M$ *and* $\eta^2 I_n \in \mathcal{C}'$ *then* $\max_{\mathcal{C}'} \mathcal{L}(C) = \mathcal{L}(\eta^2 I_n)$.

A consequence of (i) of this lemma is that if one is carrying out a statistical procedure with loss function $\mathcal{L}(C)$ then a version of the procedure that minimizes $\mathcal{L}(\eta^2 I_n)$ is minimax as C varies over $\mathcal{C}_M$. By (ii), this remains true for subsets of $\mathcal{C}_M$ that contain $\eta^2 I_n$.

An interpretation of the lemma is that, in attempting to maximize loss by altering the correlations or increasing the variances of C, one should always choose the latter. But the procedures discussed in this article do not depend on the particular value of η^2; its only role is to ensure that $\mathcal{C}_M$ is large enough to contain the departures of interest.

3. Generalized least-squares regression estimates when the response is correctly specified

Consider the linear model $y = X\theta + \varepsilon$ for $X_{n\times p}$ of rank p. Suppose that the random errors ε have covariance matrix $C \in \mathcal{C}_M$. If C is known then the best linear unbiased estimate is $\hat{\theta}_{\text{blue}} = (X^{\mathrm{T}} C^{-1} X)^{-1} X^{\mathrm{T}} C^{-1} y$. In the more common case that the covariances are at best only vaguely known, an attractive possibility is to use the generalized least squares estimate (1) for a given positive definite pseudo-precision matrix P. If $P = C^{-1}$ then the best linear unbiased estimator is returned. A diagonal P gives weighted least squares. Here we propose choosing P according to the minimax principle, i.e., to minimize the maximum value of an appropriate function $\mathcal{L}(C)$ of the covariance matrix of the estimate, as C varies over $\mathcal{C}_M$.

The covariance matrix of $\hat{\theta}_{\text{gls}}$ is

$$\operatorname{cov}(\hat{\theta}_{\text{gls}} \mid C, P) = (X^{\mathrm{T}} P X)^{-1} X^{\mathrm{T}} P C P X (X^{\mathrm{T}} P X)^{-1}.$$

Viewed as a function of C, this is nondecreasing in the Loewner ordering, so that if a function Φ is nondecreasing in this ordering then $\mathcal{L}(C \mid P) = \Phi\{\operatorname{cov}(\hat{\theta}_{\text{gls}} \mid C, P)\}$ is also nondecreasing and the conclusions of the lemma hold:

$$\max_{\mathcal{C}_M} \mathcal{L}(C \mid P) = \mathcal{L}(\eta^2 I_n \mid P) = \Phi\{\eta^2 (X^{\mathrm{T}} P X)^{-1} X P^2 X (X^{\mathrm{T}} P X)^{-1}\}.$$

But, by virtue of the Gauss–Markov theorem and the monotonicity of Φ, this last expression is minimized by $P = I_n$, i.e., by the OLS estimate $\hat{\theta}_{\text{ols}} = (X^{\mathrm{T}} X)^{-1} X^{\mathrm{T}} y$, with minimum value

$$\max_{\mathcal{C}_M} \mathcal{L}(C \mid I_n) = \Phi\{\eta^2 (X^{\mathrm{T}} X)^{-1}\}. \qquad (2)$$

It is well known that if $0 \preceq \Sigma_1 \preceq \Sigma_2$ then the ith largest eigenvalue λ_i of Σ_2 dominates that of Σ_1 for all i. It follows that Φ is nondecreasing in the Loewner ordering in the following cases:

(i) $\Phi(\Sigma) = \operatorname{tr}(\Sigma) = \sum_i \lambda_i(\Sigma)$;
(ii) $\Phi(\Sigma) = \det(\Sigma) = \prod_i \lambda_i(\Sigma)$;
(iii) $\Phi(\Sigma) = \max_i \lambda_i(\Sigma)$;
(iv) $\Phi(\Sigma) = \operatorname{tr}(K\Sigma)$ for $K \succeq 0$.

Thus, if loss is measured in any of these ways and $C \in \mathcal{C}_M$ then $\hat{\theta}_{\text{ols}}$ is minimax for $\mathcal{C}_M$ in the class of GLS estimates.

Minimax procedures are sometimes criticized for dealing optimally with an overly pessimistic least favourable case; see Huber (1972) for a discussion. Such criticism certainly does not apply here.

Remark 1. It is of interest to compare (2) to the maximum loss of the GLS estimate that assumes that the covariance is $C_0 \neq I_n$ and takes $P_0 = C_0^{-1}$. The ratio of the two maximum losses, i.e., that of the GLS estimate to that of the OLS estimate, is $\max_{\mathcal{C}_M} \mathcal{L}(C \mid P_0)/\max_{\mathcal{C}_M} \mathcal{L}(C \mid I_n)$, with each maximum attained at $\eta^2 I_n$. This is of course greater than or equal to 1, but can be arbitrarily large. See the Supplementary Material for a simple example, with C_0 as in Example 2 below and $\Phi(\Sigma) = \det(\Sigma)$, in which this ratio is unbounded.

In each of the following examples, we posit a particular covariance structure for C, a norm $\|C\|_M$, a bound η^2 and a class $\mathcal{C}'$ for which $C \in \mathcal{C}' \subseteq \mathcal{C}_M$. In each case $\eta^2 I_n \in \mathcal{C}'$, so that Lemma 1(ii) applies and $\hat{\theta}_{\text{ols}}$ is minimax for $\mathcal{C}'$ with respect to any of the criteria (i)–(iv).

Example 1 *(Independent, heteroscedastic errors).* Suppose that $C = \text{diag}(\sigma_1^2, \ldots, \sigma_n^2)$. Then the discussion above applies if $\mathcal{C}'$ is the subclass of diagonal members of $\mathcal{C}_E$ for $\eta^2 = \max_i \sigma_i^2$.

Example 2 (*Equicorrelated errors*). Suppose that the researcher fears that the observations are possibly equally correlated, and so considers $C = \sigma^2\{(1-\rho)I_n + \rho 1_n 1_n^{\mathrm{T}}\}$, with $|\rho| \leqslant \rho_{\max}$. If $\rho \geqslant 0$ then $\|C\|_1 = \|C\|_E = \sigma^2\{1 + (n-1)|\rho|\}$, and we take $\eta^2 = \sigma^2\{1 + (n-1)\rho_{\max}\}$. If $\mathcal{C}'$ is the subclass of $\mathcal{C}_1$ or $\mathcal{C}_E$ defined by the equicorrelation structure, then minimaxity of $\hat{\theta}_{\text{ols}}$ for either of these classes follows. If $\rho < 0$ then this continues to hold for $\mathcal{C}_1$, and for $\mathcal{C}_E$ if $\eta^2 = \sigma^2(1 + \rho_{\max})$.

Example 3 (*First-order moving average* MA(1) *errors*). Assume first that the random errors are homoscedastic but are possibly serially correlated, following an MA(1) model with $\text{corr}(\varepsilon_i, \varepsilon_j) = \rho I(|i-j| = 1)$ and with $|\rho| \leqslant \rho_{\max}$. Then $\|C\|_1 \leqslant \sigma^2(1 + 2\rho_{\max}) = \eta^2$, and in the discussion above we may take $\mathcal{C}'$ to be the subclass, containing $\eta^2 I_n$, defined by $c_{ij} = 0$ if $|i-j| > 1$. If the errors are instead heteroscedastic then σ^2 is replaced by $\max_i \sigma_i^2$.

Example 4 (AR(1) *errors*). It is known (see, for instance, Trench, 1999, p. 182) that the eigenvalues of an AR(1) autocorrelation matrix with autocorrelation parameter ρ are bounded, and that the maximum eigenvalue $\lambda(\rho)$ has $\lambda^* = \max_\rho \lambda(\rho) > \lambda(0) = 1$. Then, again under homoscedasticity, the covariance matrix C has $\|C\|_E \leqslant \sigma^2 \lambda^* = \eta^2$, and the discussion above applies when $\mathcal{C}'$ is the subclass defined by the autocorrelation structure.

Example 5 (*All of the above*). If $\mathcal{C}$ is the union of the classes of covariance structures employed in Examples 1–4 then the maximum loss over $\mathcal{C}$ is attained at $\eta_0^2 I_n$, where η_0^2 is the maximum of those in these four examples. Then $\hat{\theta}_{\text{ols}}$ is minimax robust against the union of these classes, since $\eta_0^2 I_n$ is in each of them.

4. MINIMAX PRECISION MATRICES IN MISSPECIFIED RESPONSE MODELS

A possibly misspecified regression response on a design space $\chi = \{x_1, \ldots, x_N\}$ is

$$E[Y|x] = f^{\mathrm{T}}(x)\theta_0 + \psi(x) \tag{3}$$

for ψ ranging over a class Ψ defined by the orthogonality constraint, $\sum_{x\in\chi} f(x)\,\psi(x) = 0_{p\times 1}$, and the boundedness constraint, $\sum_{x\in\chi} \psi^2(x) \leqslant \tau^2$ for some τ^2. The orthogonality constraint is equivalent to the definition $\theta_0 = \arg\min_\theta \sum_{x\in\chi}\{E[Y \mid x] - f^{\mathrm{T}}(x)\theta\}^2$. We consider designs $\xi = \{\xi_1, \ldots, \xi_N\}$, with $n\xi_i$ the number, perhaps zero, of observations to be placed at x_i. With $Y(x)$ drawn from (3) and $\hat{Y}(x) = f^{\mathrm{T}}(x)\hat{\theta}_{\text{gls}}(x)$, our aim is to minimize the IMSPE

$$\mathcal{I}(\xi, P \mid \psi, C) = \sum_{x\in\chi} E[\hat{Y}(x) - E\{Y(x)\}]^2,$$

after maximizing over $\psi \in \Psi$ and $C \in \mathcal{C}_M$. To express this problem in terms of the design, define a set of $n \times N$ indicator matrices

$$\mathcal{J} = \{J \in \{0,1\}^{n\times N} \mid J^{\mathrm{T}}J \text{ is diagonal with trace } n\}.$$

There is a one-one correspondence between $\mathcal{J}$ and the set of n-point designs on χ. Given J, with $J^{\mathrm{T}}J \equiv D = \mathrm{diag}(n_1, \ldots, n_N)$, the jth column of J contains n_j ones, specifying the number of occurrences of x_j in a design, which thus has design vector $\xi = n^{-1}J^{\mathrm{T}}1_n = (n_1/n, \ldots, n_N/n)^{\mathrm{T}}$. Conversely, a design ξ determines J by $J_{ij} = I\{$ the ith row of the design matrix X is $f^{\mathrm{T}}(x_j)\}$. The rank q of D is the number of support points of the design, assumed to be greater than or equal to p. Define $F_{N\times p}$ to be the matrix with rows $\{f^{\mathrm{T}}(x_i)\}_{i=1}^{N}$; then $X = JF$.

The maximum IMSPE is given in the following theorem, whose proof is provided in the Supplementary Material.

THEOREM 1. *Define* $\nu = \tau^2/(\tau^2+\eta^2)$ *; this is the relative importance to the investigator of errors due to bias rather than to variation. Then, for a design* ξ *and precision matrix* P*, the maximum of* $\mathcal{I}(\xi, P \mid \psi, C)$ *as* ψ *varies over* Ψ *and* C *varies over* $\mathcal{C}_M$ *is* $(\tau^2+\eta^2)$ *times*

$$\mathcal{I}_\nu(\xi, P) = (1-\nu)\mathcal{I}_0(\xi, P) + \nu\mathcal{I}_1(\xi, P), \tag{4}$$

where, with $U_{N\times N} = J^{\mathrm{T}}PJ$ and $V_{N\times N} = J^{\mathrm{T}}P^2J$,

$$\mathcal{I}_0(\xi, P) = \mathrm{tr}\{(Q^{\mathrm{T}}UQ)^{-1}(Q^{\mathrm{T}}VQ)(Q^{\mathrm{T}}UQ)^{-1}\},$$
$$\mathcal{I}_1(\xi, P) = \mathrm{ch}_{\max}\{(Q^{\mathrm{T}}UQ)^{-1}Q^{\mathrm{T}}U^2Q(Q^{\mathrm{T}}UQ)^{-1}\}.$$

Here the columns of $Q_{N\times p}$ *form an orthogonal basis for the column space* col(F)*,* J *is the indicator matrix of design* ξ *and* $\mathrm{ch}_{\max}$ *denotes the maximum eigenvalue of a matrix.*

To minimize $\mathcal{I}_\nu(\xi, P)$ over P for a fixed design, we invoke the Choleski decomposition to write $P = LL^{\mathrm{T}}$ for a lower triangular L. We then express $\mathcal{I}_\nu(\xi, LL^{\mathrm{T}})$ as a function of the vector $l_{n(n+1)/2\times 1}$ consisting of the elements in the lower triangle of L, and minimize over l using a nonlinear constrained minimizer. The constraint (possible since P can be multiplied by any positive scalar) is that $\mathrm{tr}P = l^{\mathrm{T}}l = n$.

The numerical results give a negative answer to the question of whether or not OLS is necessarily minimax; the minimizing P is often, but not always, the identity matrix.

In our simulation study we set the design space to be $\chi = \{-1 = x_1 < \cdots < x_N = 1\}$, with the x_i equally spaced. We chose regressors $f(x) = (1, x)^{\mathrm{T}}$, $(1, x, x^2)^{\mathrm{T}}$ or $(1, x, x^2, x^3)^{\mathrm{T}}$, corresponding to linear, quadratic or cubic regression. For various values of n and N, we first randomly generated probability distributions $(p_1, p_2, \ldots, p_N)$ and then generated a $\mathrm{Mu}(n; p_1, p_2, \ldots, p_N)$ vector; this is $n\xi$. For each such design, we computed the minimizing P, and both components of the minimized value of $\mathcal{I}_\nu(\xi, P)$. This was done for $\nu = 0.5, 1$. We took n equal to five times the number of regression parameters.

Denote by P^ν the minimizing P. Of course, $P^0 = I_n$. In each case we compared three quantities:

$$T_1 = 100\frac{\{\mathcal{I}_\nu(\xi, P^0) - \mathcal{I}_\nu(\xi, P^\nu)\}}{\mathcal{I}_\nu(\xi, P^0)} \quad \text{(the percent reduction in } \mathcal{I}_\nu \text{ achieved by} P^\nu\text{)},$$

$$T_2 = 100\frac{\{\mathcal{I}_0(\xi, P^\nu) - \mathcal{I}_0(\xi, P^0)\}}{\mathcal{I}_0(\xi, P^0)} \quad \text{(the percent increase, relative to OLS, in var)},$$

$$T_3 = 100\frac{\{\mathcal{I}_1(\xi, P^0) - \mathcal{I}_1(\xi, P^\nu)\}}{\mathcal{I}_1(\xi, P^0)} \quad \text{(the percent decrease, relative to OLS, in bias)}.$$

The means and standard errors based on 500 runs of the performance measures using these ‘multinomial’ designs are given in Table 1 in the Supplementary Material. The percentages of times that $P^\nu = I_n$ was minimax are also given. When $\nu = 1$, the percent reduction in the bias (T_3) can be significant, but is accompanied by an often sizeable increase in the variance (T_2). When $\nu = 0.5$, the reduction T_1 is typically quite modest.

These multinomial designs, mimicking those that might arise in observational studies, are not required to be symmetric. We reran the simulations after symmetrizing the designs by averaging them with their reflections across $x = 0$ and then applying a rounding mechanism that preserved symmetry. The resulting designs gave substantially reduced losses both for $P = I_n$ (OLS) and $P = P^{\nu}$ (GLS), and were much more likely to be optimized by $P^{\nu} = I_n$. The differences between the means of $\mathcal{I}_\nu(\xi, I_n)$ and $\mathcal{I}_\nu(\xi, P^{\nu})$ were generally statistically insignificant, and the values of T_1, T_2 and T_3 showed only very modest benefits to GLS. See Table 2 in the Supplementary Material.

Remark 2. From the simulations, a user might conclude that even when OLS is not minimax, the benefits of using GLS with the minimax P are outweighed by the computational complexity. An investigator who does decide beforehand to use OLS might then design to minimize $\mathcal{I}_\nu(\xi, P) = \mathcal{I}_\nu(\xi, I_n)$. This is a well-studied problem, solved for numerous response models under the assumption of independent and identically distributed errors; see Wiens (2015) for a review. We now see that these designs enjoy the additional property of being minimax against departures $C \in \mathcal{C}_M$.

Remark 3. There is an attractive class of designs for which we have proven that $\mathcal{I}_\nu(\xi, P)$ is necessarily minimized by $P = I_n$, so that OLS is minimax even in the presence of model misspecification. These are designs that are uniform on their support, as defined in § 1.

5. MINIMAX PRECISION MATRICES AND MINIMAX DESIGNS

We investigated the interplay between minimax precision matrices and minimax designs. To this end, (4) was minimized over both ξ and P . To minimize over ξ for given ν, we employed particle swarm optimization (Kennedy & Eberhart, 1995). The algorithm searches over continuous designs ξ, and so each such design to be evaluated was first rounded so that $n\xi$ had integer values. Then J and the corresponding minimax precision matrix $P^{\nu} = P^{\nu}(J)$ were computed and the loss returned. The final output is an optimal pair $\{J^{\nu}, P^{\nu}\}$. Using a genetic algorithm yielded the same results, but was many times slower.

The results, using the same responses and settings as in Tables 1 and 2 in the Supplementary Material, are shown in Table 3 in the Supplementary Material. In all cases the use of the minimax design gives significantly smaller losses, both using OLS and GLS. In eight of the 12 cases studied it turns out that the minimax design is uniform on its support and so the choice $P^{\nu} = I_n$ is minimax. In the remaining cases minimax precision results in only a marginal improvement. Of the two factors, ξ and P, explaining the decrease in $\mathcal{I}_\nu$, the design is by far the greater contributor.

See Fig. 1 for some representative plots of the minimax designs for a cubic response. For $N = 51$, the designs are uniform on their support and OLS is minimax. This fails when $N = 11$, with $T_1 = 1.5\%$ when $\nu = 0.5$ and $= 11\%$ when $\nu = 1$.

Several common features of robust designs are reflected in Fig. 1. One is that the designs using $\nu = 1$, i.e., aimed at minimization of the bias alone, tend to be more uniform than those using $\nu = 0.5$. This reflects the fact that, when a uniform design on all of χ is implemented, then the bias using OLS vanishes. As well, when the design space is sufficiently rich as to allow for clusters of nearby design points to replace replicates, then this invariably occurs. See, Fang & Wiens (2000), Heo et al. (2001) and Wiens (2024a) for examples and discussions. Such clusters form near the support points of the classically I-optimal designs, minimizing integrated variance alone. See, for instance, Studden (1977), who showed that the I-optimal design for cubic regression places masses 0.1545, 0.3455 at each of ± 1, ± 0.477, a situation well approximated by the design in Fig 1(c), whose clusters around these points account for masses of 0.15 and 0.35 each. As ν increases, the clusters become more diffuse, as in Fig. 1(d). A result of our findings in this article is that an additional benefit to such clustering is that it allows OLS to be a minimax GLS procedure.

6. CONTINUOUS DESIGN SPACES

If the design space is continuous, for instance an interval, or hypercube, then the problem of finding a minimax design is somewhat more complicated. We continue to work with the alternate

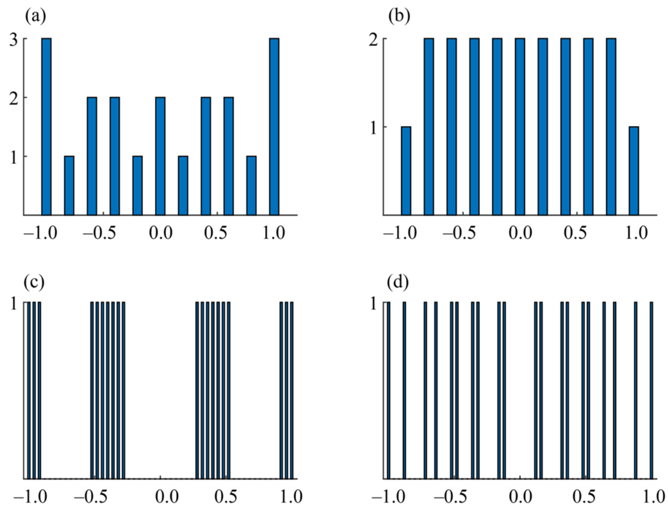

Fig. 1. Minimax design frequencies for a cubic model with $n = 20$ and (a) $N = 11, \nu = 0.5$; (b) $N = 11, \nu = 1$; (c) $N = 51, \nu = 0.5$ and (d) $N = 51, \nu = 1$.

models (3), but the orthogonality and boundedness constraints now have their finite sums replaced by Lebesgue integrals over χ. As shown by Wiens (1992), in order that a design ξ has finite maximum loss, it is necessary that it be absolutely continuous, i.e., possess a density. Otherwise, it places positive mass on some set of Lebesgue measure zero, on which any $\psi(x)$ can be modified without affecting these integrals. Such modifications can be done in such a way as to make the integrated squared bias unbounded. Thus, static, discrete designs are not admissible.

A remedy, detailed by Waite & Woods (2022), is to choose design points randomly, from an appropriate density. In the parlance of game theory, this precludes Nature, assumed malevolent, from knowing the support points of ξ and replying with a $\psi(\cdot)$ modified as above. They also recommended designs concentrated near the I-optimal design points, leading to Fig. 1(c) and (d), but with randomly chosen clusters. The resulting designs are always uniform on their supports, and so OLS is minimax in all cases, in contrast to the situation illustrated in Fig. 1(a) and (b). See Wiens (2024a) for guidelines and further examples, and Waite (2024) for extensions allowing for randomized replication. For results on robustness of inferences, see Zhang et al. (2025).

Acknowledgement

This work was supported by the Natural Sciences and Engineering Research Council of Canada.

Supplementary material

The Supplementary Material contains the proofs and other mathematical details for this article, as well as the tables related to the simulation studies. The computations were carried out in MATLAB® and the code is available on the author's personal website.

References

Aitken, A. C. (1935). On least squares and linear combinations of observations. *Proc. R. Soc. Edin.* **55**, 42–8.

Cochrane, D. & Orcutt, G. H. (1949). Application of least squares regression to relationships containing autocorrelated error terms. *J. Am. Statist. Assoc.* **44**, 32–61.

Fang, Z. & Wiens, D. P. (2000). Integer-valued, minimax robust designs for estimation and extrapolation in heteroscedastic, approximately linear models. *J. Am. Statist. Assoc.* **95**, 807–18.

Fomby, T. B., Johnson, S. R. & Hill, R. C. (1984). Feasible generalized least squares estimation. In *Advanced Econometric Methods*, pp. 147–69. New York: Springer.

HEO, G., SCHMULAND, B. & WIENS, D. P. (2001). Restricted minimax robust designs for misspecified regression models. *Can. J. Statist*. **29**, 117–28.

HUBER, P. J. (1972). The 1972 Wald lecture robust statistics: a review. *Ann. Math. Statist.* **43**, 1041–67.

KENNEDY, J. & EBERHART, R. (1995). Particle swarm optimization. In *Proc. ICNN'95 Int. Conf. Neural Networks*, pp. 1942–8. Piscataway, NJ: IEEE Press.

STUDDEN, W. J. (1977). Optimal designs for integrated variance in polynomial regression. In *Statistical Decision Theory and Related Topics, Ed. S. S. Gupta and D. S. Moore*, pp. 411–20. New York: Academic Press.

TRENCH, W. F. (1999). Asymptotic distribution of the spectra of a class of generalized Kac-Murdock-Szegö matrices. *Lin. Algeb. Applic*. **294**, 181–92.

WAITE, T. W. (2024). Replication in random translation designs. *Statist. Prob. Lett*. **215**, 110229.

WAITE, T. W. & WOODS, D. C. (2022). Minimax efficient random experimental design strategies with application to model-robust design for prediction. *J. Am. Statist. Assoc*. **117**, 1452–65.

WIENS, D. P. (1992). Minimax designs for approximately linear regression. *J. Statist. Plan. Infer*. **31**, 353–71.

WIENS, D. P. (2015). Robustness of design. In *Handbook of Design and Analysis of Experiments, Ed.* A. Dean, M. Morris, J. Stufken and D. Bingham, pp. 719–53. New York: Chapman and Hall/CRC Press.

WIENS, D. P. (2024a). Jittering and clustering: strategies for the construction of robust designs. *Statist. Comp*. **34**, 129.

WIENS, D. P. (2024b). A note on minimax robustness of designs against correlated or heteroscedastic responses. *Biometrika* **111**, 1071–5.

ZHANG, Z., DING, P., ZHOU, W. & WANG, H. (2025). With random regressors, least squares inference is robust to correlated errors with unknown correlation structure. *Biometrika* **112**, asae054.

[*Received on* 7 *June* 2024. *Editorial decision on* 31 *March* 2025]