To ignore dependencies is perhaps not a sin

Douglas P. Wiens[a,*]

[a]*Mathematical & Statistical Sciences, University of Alberta, Edmonton, Canada, T6G 2G1*

August 9, 2024

Abstract

We present 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 i.i.d. data can be minimax. In particular, the ordinary least squares (OLS) estimate of a correctly specified regression response is minimax among generalized least squares (GLS) estimates, when the maximum is taken over certain classes of error covariance structures and the loss function possesses a natural monotonicity property. An implication is that it can be not only safe, but optimal to ignore such departures from the usual assumption of i.i.d. errors. We then consider regression models in which the response function is possibly misspecified, and show that OLS is minimax if the design is uniform on its support, but that this often fails otherwise. We go on to investigate the interplay between minimax GLS procedures and minimax designs, leading 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.

Keywords: design, induced matrix norm, Loewner ordering, particle swarm optimization, robustness.
2010 MSC: Primary 62G35, Secondary 62K05

1. Introduction and summary

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 and, even if there is some prior knowledge of its structure, before the study is carried out there are no data from which accurate estimates of its elements might be made. If a consistent estimate $\hat{C}^{-1}$ of the precision matrix does exist, then one can employ 'feasible generalized least squares' estimation - see e.g. Fomby et al. (1984). An example is the Cochrane-Orcutt procedure (Cochrane and Orcutt (1949)), which can be applied iteratively in 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

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

In Wiens (2024b) a similar problem was addressed, pertaining to designed experiments whose data are to be analyzed by OLS. A lemma, restated below as Lemma 1, was used 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 independent 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. 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 regression 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 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 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 attained by replacing the replicates that are a feature of 'classically optimal' designs minimizing variance alone by clusters of observations at nearby design points.

    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.

The computations for this article were carried out in MATLAB; the code is available on the author's personal website.

2. A useful lemma

Suppose that $\|\cdot\|_M$ is a matrix norm, induced by the vector norm $\|\cdot\|_V$, i.e.

$$\|C\|_M = \sup_{\|x\|_V = 1} \|Cx\|_V.$$

We use the subscript 'M' when referring to an arbitrary matrix norm, but adopt special notation in the following cases:

(i) For the Euclidean norm $\|x\|_V = (x'x)^{1/2}$, the matrix norm is denoted $\|C\|_E$ and is the spectral radius, i.e. the root of the maximum eigenvalue of $C'C$. This is the maximum eigenvalue of C if C is a covariance matrix, i.e. is symmetric and positive semidefinite.

(ii) For the sup norm $\|x\|_V = \max_i |x_i|$, the matrix norm $\|C\|_\infty$ is $\max_i \sum_j |c_{ij}|$, the maximum absolute row sum.

(iii) For the 1-norm $\|x\|_V = \sum_i |x_i|$, the matrix norm $\|C\|_1$ is $\max_j \sum_i |c_{ij}|$, the maximum absolute column sum. For symmetric matrices, $\|C\|_1 = \|C\|_\infty$.

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 non-decreasing in the Loewner ordering:

$$A \preceq B \Rightarrow \mathcal{L}(A) \leq \mathcal{L}(B).$$

Here $A \preceq B$ means that $B - A \succeq 0$, i.e. is positive semidefinite (p.s.d.).

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

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

$$C_M = \left\{ C \mid C \succeq 0 \text{ and } \|C\|_M \leq \eta^2 \right\}. \tag{2}$$

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

$$C_E = \left\{ C \mid 0 \preceq C \preceq \eta^2 I_n \right\}.$$

Then:
(i) In any such class C_M, $\max_{C_M} \mathcal{L}(C) = \mathcal{L}\left(\eta^2 I_n\right)$.
(ii) If $C' \subseteq C_M$ and $\eta^2 I_n \in C'$, then $\max_{C'} \mathcal{L}(C) = \mathcal{L}\left(\eta^2 I_n\right)$.

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 which minimizes $\mathcal{L}\left(\eta^2 I_n\right)$ is *minimax* as C varies over C_M.

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 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

$$y = X\theta + \varepsilon \tag{3}$$

for $X_{n \times p}$ of rank p. Suppose that the random errors ε have covariance matrix $C \in C_M$. If C is *known* then the 'best linear unbiased estimate' is $\hat{\theta}_{\text{BLUE}} = \left(X'C^{-1}X\right)^{-1}X'C^{-1}y$. 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 BLUE is returned. A diagonal P gives 'weighted least squares' (WLS). 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 C_M.

For brevity we drop the 'pseudo' and call P a precision matrix. Since $\hat{\theta}_{\text{GLS}}$ is invariant under multiplication of P by a scalar, we assume throughout that

$$tr(P) = n. \tag{4}$$

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

$$\text{cov}\left(\hat{\theta}_{\text{GLS}} \mid C, P\right) = (X'PX)^{-1}X'PCPX(X'PX)^{-1}.$$

Viewed as a function of C this is non-decreasing in the Loewner ordering, so that if a function Φ is non-decreasing in this ordering, then

$$\mathcal{L}(C \mid P) = \Phi\{\text{cov}(\hat{\theta}_{\text{GLS}} \mid C, P)\}$$

is also non-decreasing and the conclusions of the lemma hold:

$$\max_{C_M} \mathcal{L}(C \mid P) = \mathcal{L}\left(\eta^2 I_n \mid P\right) = \Phi\left\{\eta^2 (X'PX)^{-1}XP^2X(X'PX)^{-1}\right\}.$$

But this last expression is minimized by $P = I_n$, i.e. by the OLS estimate $\hat{\theta}_{\text{OLS}} = (X'X)^{-1}X'y$, with minimum value

$$\max_{C_M} \mathcal{L}(C \mid I_n) = \Phi\left\{\eta^2 (X'X)^{-1}\right\}.$$

This follows from the monotonicity of Φ and the inequality

$$\eta^2 (X'PX)^{-1}X'P^2X(X'PX)^{-1} - \eta^2 (X'X)^{-1}$$
$$= \eta^2 (X'PX)^{-1}X'P\left\{I_n - X(X'X)^{-1}X'\right\}PX(X'PX)^{-1} \geq 0,$$

which uses the fact that $I_n - X(X'X)^{-1}X'$ is idempotent, hence positive semidefinite.

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

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

Thus if loss is measured in any of these ways and $C \in C_M$ then $\hat{\theta}_{\mathrm{OLS}}$ is minimax for 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 does certainly not apply here.

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 C' for which $C \in C' \subseteq C_M$. In each case $\eta^2 I_n \in C'$, so that statement (ii) of the lemma applies and $\hat{\theta}_{\mathrm{OLS}}$ is minimax for C' (and for all of C_M as well) and with respect to any of the criteria (i) – (iv).

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

Example 2: Equicorrelated errors. Suppose that the researcher fears that the observations are possibly weakly correlated, and so considers $C = \sigma^2 ((1-\rho) I_n + \rho 1_n 1_n')$, with $|\rho| \leq \rho_{\max}$. If $\rho \geq 0$ then $\|C\|_1 = \|C\|_\infty = \|C\|_E = \sigma^2 \{1 + (n-1)|\rho|\}$, and we take $\eta^2 = \sigma^2 \{1 + (n-1)\rho_{\max}\}$. If C' is the subclass of C_1 or C_∞ or C_E defined by the equicorrelation structure, then minimaxity of $\hat{\theta}_{\mathrm{OLS}}$ for any of these classes follows. If $\rho < 0$ then this continues to hold for $C_1 = C_\infty$, and for C_E if $\eta^2 = \sigma^2 (1 + \rho_{\max})$.

Example 3: MA(1) errors. Assume first that the random errors are homoscedastic but are possibly serially correlated, following an MA(1) model with $\mathrm{CORR}(\varepsilon_i, \varepsilon_j) = \rho I (|i - j| = 1)$ and with $|\rho| \leq \rho_{\max}$. Then $\|C\|_1 = \|C\|_\infty \leq \sigma^2 (1 + 2\rho_{\max}) = \eta^2$, and in the discussion above we may take 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 $\|C\|_E \leq \sigma^2 \lambda^* = \eta^2$, and the discussion above applies when C' is the subclass defined by the autocorrelation structure.

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

3.1. Inference from GLS estimates when $C = \sigma^2 I_n$

In the next section we consider biased regression models, and investigate the performance of the GLS estimate (1) with $P \neq I_n$ even though C is a multiple σ^2 of the identity matrix. A caveat to the use of this estimate in correctly specified models (3) is given by

the following result. It was established in Wiens (2000) for wls estimates, but holds for gls estimates as well.

Theorem 1. *Suppose that data $y_{n\times 1}$ obey the linear model (3) with $C = \sigma^2 I_n$ and that a gls estimate (1), with $P \neq I_n$, is employed. Let $H : n \times n$ be the projector onto the column space of $(X \vdots PX) : n \times 2p$. Then an unbiased estimate of σ^2 is*

$$S^2 = \|(I_n - H)\,y\|^2 \,/\, (n - rk\,(H))\,.$$

The vector $(I_n - H)\,y$ is uncorrelated with $\hat{\theta}_{\text{GLS}}$. If the errors are normally distributed, then $S^2 \sim \sigma^2 \chi^2_{n-rk(H)}$, independently of $\hat{\theta}_{\text{GLS}}$.

The projector H will typically have rank $2p$ when $P \neq I_n$, and so p degrees of freedom are lost in the estimation of σ^2 and subsequent normal-theory inferences.

4. Minimax precision matrices in misspecified response models

Working in finite design spaces $\chi = \{x_i\}_{i=1}^N \subset \mathbb{R}^d$, and with p-dimensional regressors $f(x)$, Wiens (2018) studied design problems for possibly misspecified regression models

$$Y(x) = f'(x)\,\theta + \psi(x) + \varepsilon, \tag{5}$$

with the unknown contaminant ψ ranging over a class Ψ and satisfying, for identifiability of θ, the orthogonality condition

$$\sum_{x\in\chi} f(x)\,\psi(x) = 0_{p\times 1}, \tag{6}$$

as well as a bound

$$\sum_{x\in\chi} \psi^2(x) \le \tau^2. \tag{7}$$

For designs ξ placing mass ξ_i on $x_i \in \chi$, he took $\hat{\theta} = \hat{\theta}_{\text{OLS}}$, loss function imspe:

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

and found designs minimizing the maximum, over ψ, of $\mathcal{I}(\psi,\xi)$.

In Wiens (2018) the random errors ε_i were assumed to be i.i.d.; now suppose that they instead have covariance matrix $C \in C_M$ and take $\hat{\theta} = \hat{\theta}_{\text{GLS}}$ with precision matrix P. Using (6), and emphasizing the dependence on C and P, $\mathcal{I}(\psi,\xi)$ decomposes as

$$\mathcal{I}(\psi,\xi\,|\,C,P) = \sum_{x\in\chi} f'(x)\,\text{cov}\left(\hat{\theta}\,|\,C,P\right)f(x) + \sum_{x\in\chi} f'(x)\,b_{\psi,P}b'_{\psi,P}f(x) + \sum_{x\in\chi} \psi^2(x)\,. \tag{8}$$

Here $b_{\psi,P} = E(\hat{\theta}) - \theta$ is the bias. Denote by ψ_X the $n \times 1$ vector consisting of the values of ψ corresponding to the rows of X, so that $b_{\psi,P} = (X'PX)^{-1}\,X'P\psi_X$.

To express these quantities in terms of the design, define a set of $n \times N$ indicator matrices

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

There is a one-one correspondence between $\mathcal{J}$ and the set of n-point designs on χ. Given J, with

$$J'J \equiv D = diag\,(n_1, ..., n_N),$$

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 $\xi = n^{-1}J'1_n = (n_1/n, ..., n_N/n)'$. Conversely, a design determines J by $J_{ij} = I\big(\text{the } i\text{th row of } X \text{ is } f'\,(x_j)\big)$. The rank q of D is the number of support points of the design, assumed $\geq p$.

Define $F_{N \times p}$ to be the matrix with rows $\{f'\,(x_i)\}_{i=1}^N$. Then $X = JF$ and, correspondingly, $\psi_X = J\bar{\psi}$ for $\bar{\psi} = (\psi\,(x_1), ..., \psi\,(x_N))'$.

The proofs for this section are in the appendix.

Theorem 2. *For η as at (2) and τ as at (7), define $v = \tau^2 / \big(\tau^2 + \eta^2\big)$; 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 $\mathcal{I}\,(\psi, \xi \mid C, P)$ as C varies over C_M and ψ varies over Ψ is $\big(\tau^2 + \eta^2\big) \times$*

$$\mathcal{I}_v\,(\xi, P) = (1 - v)\,\mathcal{I}_0\,(\xi, P) + v\mathcal{I}_1\,(\xi, P), \tag{9}$$

where

$$\mathcal{I}_0\,(\xi, P) = tr\left\{(Q'UQ)^{-1}\,(Q'VQ)\,(Q'UQ)^{-1}\right\}, \tag{10a}$$

$$\mathcal{I}_1\,(\xi, P) = ch_{\max}\left\{(Q'UQ)^{-1}\,Q'U^2Q\,(Q'UQ)^{-1}\right\}, \tag{10b}$$

$$U_{N \times N} = J'PJ \text{ and } V_{N \times N} = J'P^2J.$$

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

Remark 1. *We assume throughout that the design is such that $X'PX > 0$, implying that $Q'UQ > 0$.*

Remark 2. *An investigator might decide beforehand to use OLS, and then design to minimize $\mathcal{I}_v\,(\xi, P) = \mathcal{I}_v\,(\xi, I_n)$. This is a well-studied problem, solved for numerous response models under the assumption of i.i.d. errors – see Wiens (2015) for a review. By virtue of Theorem 2 these designs enjoy the additional property of being minimax against departures $C \in C_M$.*

4.1. Simulations

In (9), VAR $= \mathcal{I}_0$ and BIAS $= \mathcal{I}_1$ are the components of the IMSPE due to variation and to bias, respectively. That $\mathcal{I}_0\,(\xi, P)$ is minimized by OLS for any design was established in §3. If OLS is to be a minimax procedure for a particular design and some $v \in (0, 1]$, any

Table 1. Minimax precision matrices; multinomial designs:
means of performance measures ± 1 standard error.

Response	N	v	$\%I_n$	$\bar{\mathcal{I}}_v(\xi, I_n)$	$\bar{\mathcal{I}}_v(\xi, P^v)$	$T_1\,(\%)$	$T_2\,(\%)$	$T_3\,(\%)$
	11	.5	1	3.34 ± .11	3.19 ± .11	4.16 ± .15	3.23 ± .12	12.29 ± .46
linear	11	1	1	3.72 ± .16	3.27 ± .14	11.81 ± .35	9.18 ± .31	14.40 ± .52
$n = 10$	51	.5	27	11.19 ± .21	11.05 ± .21	1.24 ± .07	.85 ± .05	4.10 ± .22
	51	1	27	9.80 ± .23	9.35 ± .22	4.34 ± .22	2.96 ± .16	4.82 ± .26
	11	.5	0	5.99 ± 1.25	5.57 ± 1.06	4.62 ± .15	3.15 ± .11	14.16 ± .50
quadratic	11	1	0	7.30 ± 1.82	6.07 ± 1.27	13.04 ± .38	8.57 ± .38	16.22 ± .57
$n = 15$	51	.5	4	12.61 ± .46	12.40 ± .45	1.58 ± .07	.99 ± .05	5.75 ± .27
	51	1	4	10.69 ± .54	10.03 ± .51	5.95 ± .24	3.54 ± .17	6.72 ± .31
	11	.5	0	9.71 ± 1.63	9.15 ± 1.52	4.87 ± .15	3.25 ± .11	14.90 ± .51
cubic	11	1	0	12.98 ± 2.53	11.41 ± 2.22	13.54 ± .38	8.86 ± .29	16.92 ± .58
$n = 20$	51	.5	0	21.67 ± 2.43	21.21 ± 2.38	1.87 ± .08	1.14 ± .05	7.24 ± .30
	51	1	0	20.76 ± 2.9	19.29 ± 2.81	7.39 ± .26	3.75 ± .16	8.45 ± .35

Table 2. Minimax precision matrices; symmetrized designs:
means of performance measures ± 1 standard error.

Response	N	v	$\%I_n$	$\bar{\mathcal{I}}_v(\xi, I_n)$	$\bar{\mathcal{I}}_v(\xi, P^v)$	$T_1\,(\%)$	$T_2\,(\%)$	$T_3\,(\%)$
	11	.5	20	2.10 ± .03	2.05 ± .03	2.45 ± .11	1.64 ± .07	8.46 ± .41
linear	11	1	20	1.83 ± .04	1.66 ± .04	8.51 ± .31	4.99 ± .21	10.06 ± .46
$n = 10$	51	.5	80	10.01 ± .29	9.97 ± .29	.40 ± .05	.23 ± .03	1.45 ± .18
	51	1	80	7.77 ± .29	7.67 ± .29	1.48 ± .16	.67 ± .07	1.66 ± .20
	11	.5	0	2.35 ± .08	2.26 ± .07	3.49 ± .07	2.17 ± .07	14.10 ± .50
quadratic	11	1	0	2.01 ± .09	1.74 ± .08	14.40 ± .37	8.57 ± .22	18.05 ± .58
$n = 15$	51	.5	85	10.58 ± .70	10.53 ± .68	.25 ± .04	.15 ± .02	1.02 ± .16
	51	1	85	7.54 ± .80	7.39 ± .74	1.03 ± .15	.49 ± .07	1.19 ± .19
	11	.5	3	2.64 ± .19	2.55 ± .19	3.56 ± .12	1.99 ± .07	15.18 ± .56
cubic	11	1	3	2.39 ± .27	2.11 ± .26	15.19 ± .44	9.09 ± .28	19.69 ± .70
$n = 20$	51	.5	58	11.97 ± 1.91	11.80 ± 1.80	.45 ± .04	.24 ± .02	2.11 ± .19
	51	1	58	8.44 ± 2.14	7.94 ± 1.80	2.23 ± .18	.86 ± .07	2.47 ± .21

increase in BIAS must be outweighed by a proportional decrease in VAR. We shall present theoretical and numerical evidence that for many designs this is not the case, and for these pairs (ξ, v) OLS is not minimax.

We first present the results of a simulation study, in which the designs exhibit no particular structure. To find P minimizing (9) we use the fact that, by virtue of the Choleski decomposition, any positive definite matrix can be represented as $P = LL'$, for a lower triangular L. We thus express $\mathcal{I}_v(\xi, LL')$ 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 – recall (4) – is that $l'l = n$.

Of course we cannot guarantee that this yields an absolute minimum, but the numerical evidence is compelling. In any event, 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 $f(x) = (1, x)'$, $\left(1, x, x^2\right)'$ or $\left(1, x, x^2, x^3\right)'$, corresponding to linear, quadratic, or cubic regression. For various values of n and N we first randomly generated probability distributions $(p_1, p_2, .., p_N)$ and then generated a multinomial$(n; p_1, p_2, .., p_N)$ vector; this is $n\xi$. For each such design we computed the minimizing P, and both components of the minimized value of $\mathcal{I}_v(\xi, P)$. This was done for $v = .5, 1$. We took n equal to five times the number of regression parameters. Denote by P^v the minimizing P. Of course $P^0 = I_n$. In each case we compared three quantities:

$$T_1 = 100\frac{\left(\mathcal{I}_v\left(\xi, P^0\right) - \mathcal{I}_v\left(\xi, P^v\right)\right)}{\mathcal{I}_v\left(\xi, P^0\right)}, \text{ the percent reduction in } \mathcal{I}_v \text{ achieved by } P^v;$$

$$T_2 = 100\frac{\left(\mathcal{I}_0\left(\xi, P^v\right) - \mathcal{I}_0\left(\xi, P^0\right)\right)}{\mathcal{I}_0\left(\xi, P^0\right)}, \text{ the percent increase, relative to OLS, in VAR;}$$

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

The means, and standard errors based on 500 runs, of the performance measures using these 'multinomial' designs are given in Table 1. The percentages of times that $P^v = I_n$ was minimax are also given. When $v = 1$ 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 $v = .5$ the reduction T_1 is typically quite modest.

These multinomial designs, mimicking those which might arise in observational studies, are not required to be symmetric. We re-ran the simulations after symmetrizing the designs by averaging them with their reflections across $x = 0$ and then applying a rounding mechanism which preserved symmetry. The resulting designs gave substantially reduced losses both for $P = I_n$ (OLS) and $P = P^v$ (GLS), and were much more likely to be optimized by $P^v = I_n$. The differences between the means of $\mathcal{I}_v(\xi, I_n)$ and $\mathcal{I}_v(\xi, P^v)$ 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.

A practitioner might understandably conclude that, even though P^ν is minimax, its benefits are outweighed by the computational complexity of its implementation. This is bolstered by Theorem 1, which continues to hold with the modification that S^2 now follows a scaled non-central χ^2 distribution, with a non-centrality parameter depending on $\psi'_X (I_n - H) \psi_X$.

4.2. Theoretical complements

In Theorem 3 below, we show that the experimenter can often design in such a way that $P = I_n$ is a minimax precision matrix, so that OLS is a minimax procedure. In particular, this holds if the design is *uniform* on its support, i.e. places an equal number of observations at each of several points of the design space.

Suppose that a design ξ places $n_i \geq 0$ observations at $x_i \in \chi$. Let $J_+ : n \times q$ be the result of retaining only the non-zero columns of J, so that $JJ' = J_+ J'_+$, and $D_+ \equiv J'_+ J_+$ is the diagonal matrix containing the positive n_i. If the columns removed have labels $j_1, ..., j_{N-q}$ then let $Q_+ : q \times p$ be the result of removing these rows from Q, so that $JQ = J_+ Q_+$ and $Q'DQ = Q'_+ D_+ Q_+$. Now define $\alpha = n/tr\left(D_+^{-1}\right)$ and

$$P_0 = \alpha J_+ D_+^{-2} J'_+, \tag{11}$$

with $trP_0 = n$. Note that

$$rk(P_0) = rk\left(J_+ D_+^{-1}\right) = rk\left(D_+^{-1} J'_+ J_+ D_+^{-1}\right) = rk\left(D_+^{-1}\right) = q,$$

so that P_0 is positive definite iff $q = n$. This is relevant in part (ii) of Theorem 3, where we deal with the possible rank deficiency of P_0 by introducing

$$P_\varepsilon \equiv (P_0 + \varepsilon I_n) / (1 + \varepsilon); \tag{12}$$

for $\varepsilon > 0$, P_ε is positive definite with $tr(P_\varepsilon) = n$.

Theorem 3. *(i) Suppose that $q \leq N$ and the design is uniform on q points of χ, with $k \geq 1$ observations at each x_i. Then $n = kq$, $D_+ = kI_q$, and $P = I_n$ is a minimax precision matrix:*

$$\mathcal{I}_\nu(\xi, I_n) = \min_{P>0} \mathcal{I}_\nu(\xi, P); \tag{13}$$

thus OLS is minimax within the class of GLS methods. In particular this holds if $P_0 = I_n$, where P_0 is defined at (11).

(ii) Suppose that a design ξ places mass on $q \leq N$ points of χ, that $P_0 \neq I_n$, and that neither of the following holds:

$$(Q'_+ Q_+)^{-1} \left(Q'_+ D_+^{-1} Q_+\right)(Q'_+ Q_+)^{-1} = (Q'_+ D_+ Q_+)^{-1}, \tag{14a}$$

$$ch_{\max}\left\{(Q'_+ D_+ Q_+)^{-1}\left(Q'_+ D_+^2 Q_+\right)(Q'_+ D_+ Q_+)^{-1}\right\} = ch_{\max}\left\{(Q'_+ Q_+)^{-1}\right\}. \tag{14b}$$

Then in particular D_+ is not a multiple of I_q and so the design is non-uniform. With P_ε as defined at (12), there is $\nu_0 \in (0, 1)$ for which, for each $\nu \in (\nu_0, 1]$, $\mathcal{I}_\nu(\xi, P_\varepsilon) < \mathcal{I}_\nu(\xi, I_n)$. Thus OLS is not minimax for such (ξ, ν).

Table 3. Minimax designs and precision matrices:
performance measures ($T_1 = T_2 = T_3 = 0$ if $P^v = I_n$).

Response	N	v	$\mathcal{I}_v(\xi, I_n)$	$\mathcal{I}_v(\xi, P^v)$	$T_1(\%)$	$T_2(\%)$	$T_3(\%)$
	11	.5	1.60	1.60	0	0	0
linear	11	1	1.10	1.10	0	0	0
$n = 10$	51	.5	6.14	6.14	0	0	0
	51	1	5.10	5.10	0	0	0
	11	.5	1.61	1.53	4.67	2.81	16.06
quadratic	11	1	1.12	1.00	10.84	10.78	10.84
$n = 15$	51	.5	5.80	5.80	0	0	0
	51	1	3.40	3.40	0	0	0
	11	.5	1.55	1.53	1.46	1.13	6.04
cubic	11	1	1.12	1.00	11.03	4.84	11.03
$n = 20$	51	.5	5.56	5.56	0	0	0
	51	1	2.55	2.55	0	0	0

Remark 3. *The requirement of Theorem 3(ii) that (14a) and (14b) fail excludes more designs than those which are uniform on their supports, and is a condition on Q as well as on the design. For instance if $Q_+ (Q'_+ Q_+)^{-1/2} \equiv A_{q \times p}$ is block-diagonal: $A = \oplus_{i=1}^{m} A_i$, where $A_i : q_i \times p_i$ ($\sum q_i = q, \sum p_i = p$) satisfies $A'_i A_i = I_{p_i}$, and if $D_+ = \oplus_{i=1}^{m} k_i I_{q_i}$, then*

$$A'D_+^{-1}A \;=\; (A'D_+A)^{-1}, \tag{15}$$

$$(A'D_+A)^{-1} A'D_+^2A (A'D_+A)^{-1} \;=\; I_p. \tag{16}$$

Equation (15) gives (14a), and (16) asserts the equality of the two matrices in (14b), hence of their maximum eigenvalues. These equations are satisfied even though the design is non-uniform if the k_i are not all equal.

In Tables 1 and 2, uniform designs account for 100% and 95%, respectively, of the cases in which $P^v = I_n$ is optimal. Common exceptions in Table 2 are designs which are uniform apart from having points added or removed at $x = 0$ to maintain symmetry. Those designs for which I_n is not optimal all meet the conditions of Theorem 3(ii). This was checked numerically: since (14a) implies that $\mathcal{I}_0(\xi, P_0) - \mathcal{I}_0(\xi, I_n) = 0$, and (14b) implies that $\mathcal{I}_1(\xi, I_n) - \mathcal{I}_1(\xi, P_0) = 0$, their failure is verified by checking that each of these differences is positive.

5. Minimax precision matrices and minimax designs

We investigated the interplay between minimax precision matrices and minimax designs. To this end (9) was minimized over both ξ and P. To minimize over ξ we employed particle swarm optimization (Kennedy and 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 $P^v = P^v(J)$
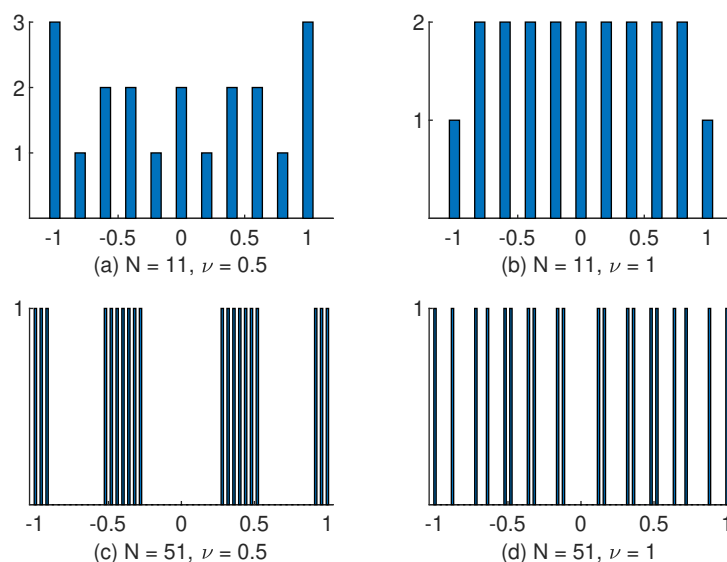
Figure 1: Minimax design frequencies for a cubic model; $n = 20$.

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 parameters as in Tables 1 and 2, are shown in Table 3. We note that in all cases the use of the minimax design gives significantly smaller losses, both using OLS and GLS. In eight of the twelve 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 – all in line with (iii) of Theorem 3 – 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 Figure 1 for some representative plots of the minimax designs for a cubic response. These reflect several common features of robust designs. 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 $\nu = .5$. This reflects the fact – following from (6) and exploited in (i) of Theorem 3 – 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 Wiens (2024a), Fang and Wiens (2000) and Heo et al. (2001) for examples and discussions. Such clusters form near the support points of the classically I-optimal designs, minimizing variance alone. See for instance Studden (1977) who showed that the I-optimal design for cubic regression places masses .1545, .3455 at each of ±1, ±.477 - a situation well approximated by the design in (c) of Figure 1, whose clusters around these points account for masses of .15 and .35 each. As ν increases the clusters become more diffuse, as in d) of Figure 1. 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 – an interval, or hypercube, for instance – then the problem of finding a minimax design is somewhat more complicated. We continue to work with the alternate models (5), but the constraints (6) and (7) now have their finite sums replaced by Lebesgue integrals over χ. As shown in Wiens (1992), in order that a design ξ have 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 and 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 recommend designs concentrated near the I-optimal design points, leading to (c) and (d) of Figure 1, 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 (a) and (b) of Figure 1. See Wiens (2024a) for guidelines and further examples, and Waite (2024) for extensions allowing for randomized replication.

Appendix: Derivations

Proof of Theorem 2. In the notation of the theorem, (8) becomes

$$
\begin{aligned}
\mathcal{I}(\psi, \xi \mid C, P) &= tr\left\{F\mathrm{cov}\left(\hat{\theta} \mid C, P\right)F'\right\} \\
&+ \bar{\psi}'J'PJF\left(F'J'PJF\right)^{-1}F'F\left(F'J'PJF\right)^{-1}F'J'PJ\bar{\psi} + \bar{\psi}'\bar{\psi}. \quad \text{(A.1)}
\end{aligned}
$$

As in §3, and taking $K = F'F$ in (iv) of that section, for $C \in \mathcal{C}_M$ the trace in (A.1) is maximized by $C = \eta^2 I_n$, with

$$
tr F\mathrm{cov}\left(\hat{\theta} \mid \eta^2 I_n, P\right)F' = \eta^2 tr\left\{F\left(F'J'PJF\right)^{-1}\left(F'J'P^2JF\right)\left(F'J'PJF\right)^{-1}F'\right\}. \quad \text{(A.2)}
$$

Extend the orthogonal basis for $col(F)$ – formed by the columns of Q – by appending to Q the matrix $Q_* : N \times (N - p)$, whose columns form an orthogonal basis for the orthogonal complement $col(F)^\perp$. Then $(Q \vdots Q_*) : N \times N$ is an orthogonal matrix and we have that $F = QR$ for a non-singular R. If the construction is carried out by the Gram-Schmidt method, then R is upper triangular.
Constraint (6) dictates that ψ lie in $col(Q_*)$. A maximizing ψ will satisfy (7) with equality, hence $\bar{\psi} = \tau Q_* \beta$ for some $\beta_{(N-p) \times 1}$ with unit norm. Combining these observations along with (A.1) and (A.2) yields that $\max_{\psi, C} \mathcal{I}(\psi, \xi \mid C, P)$ is given by

$$
\begin{aligned}
&\eta^2 tr\left\{Q\left(Q'UQ\right)^{-1}\left(Q'VQ\right)\left(Q'UQ\right)^{-1}Q'\right\} \\
&+\tau^2 \max_{\|\beta\|=1}\left\{\beta'Q_*'UQ\left(Q'UQ\right)^{-1}Q'Q\left(Q'UQ\right)^{-1}Q'UQ_*\beta + 1\right\}. \quad \text{(A.3)}
\end{aligned}
$$

Here and elsewhere we use that $trAB = trBA$, and that such products have the same non-zero eigenvalues. Then (A.3) becomes $\left(\tau^2 + \eta^2\right)$ times $\mathcal{I}_v(\xi, P)$, given by

$$\mathcal{I}_v(\xi, P) = (1-v)\, tr\left\{(Q'UQ)^{-1}(Q'VQ)(Q'UQ)^{-1}\right\}$$
$$+v\left\{ch_{\max}Q'_*UQ(Q'UQ)^{-1}\cdot(Q'UQ)^{-1}Q'UQ_* + 1\right\}. \tag{A.4}$$

The maximum eigenvalue is also that of

$$\begin{aligned}
(Q'UQ)^{-1}Q'UQ_*\cdot Q'_*UQ(Q'UQ)^{-1} &= (Q'UQ)^{-1}Q'U(I_N - QQ')UQ(Q'UQ)^{-1}\\
&= (Q'UQ)^{-1}Q'U^2Q(Q'UQ)^{-1} - I_p;
\end{aligned}$$

this in (A.4) gives (9). □

The proof of Theorem 3 requires a preliminary result.

Lemma 2. *(i) For a fixed design ξ and any $P > 0$, $\mathcal{I}_0(\xi, P) \ge \mathcal{I}_0(\xi, I_n)$ and $\mathcal{I}_1(\xi, P) \ge ch_{\max}\left\{(Q'_+Q_+)^{-1}\right\}$.*
If neither of the equations (14a), (14b) holds, then:
(ii) $\mathcal{I}_0(\xi, P_0) > \mathcal{I}_0(\xi, I_n)$ and $\mathcal{I}_1(\xi, I_n) > \mathcal{I}_1(\xi, P_0)$;
(iii) With P_ε as defined in Theorem 3(ii), and for sufficiently small $\varepsilon > 0$, $\Delta_0(\varepsilon) = \mathcal{I}_0(\xi, P_\varepsilon) - \mathcal{I}_0(\xi, I_n) > 0$ and $\Delta_1(\varepsilon) = \mathcal{I}_1(\xi, I_n) - \mathcal{I}_1(\xi, P_\varepsilon) > 0$.

Proof of Lemma 2. (i) From (10a), $\mathcal{I}_0(\xi, P) - \mathcal{I}_0(\xi, I_n)$ is the trace of

$$\begin{aligned}
&(Q'J'PJQ)^{-1}\left(Q'J'P^2JQ\right)(Q'J'PJQ)^{-1} - (Q'J'JQ)^{-1}\\
&= (Q'J'PJQ)^{-1}Q'J'P\left\{I_n - JQ(Q'J'JQ)^{-1}Q'J'\right\}PJQ(Q'J'PJQ)^{-1},
\end{aligned}$$

which is ≥ 0 (since the matrix in braces is idempotent, hence p.s.d.) with non-negative trace. For the second inequality first note that

$$\begin{aligned}
&(Q'UQ)^{-1}Q'U^2Q(Q'UQ)^{-1} - (Q'_+Q_+)^{-1}\\
&= (Q'_+J'_+PJ_+Q_+)^{-1}Q'_+J'_+PJ_+J'_+PJ_+Q_+(Q'_+J'_+PJ_+Q_+)^{-1} - (Q'_+Q_+)^{-1}\\
&= (Q'_+J'_+PJ_+Q_+)^{-1}Q'_+J'_+PJ_+\left\{I_n - Q_+(Q'_+Q_+)^{-1}Q'_+\right\}J'_+PJ_+Q_+(Q'_+J'_+PJ_+Q_+)^{-1}
\end{aligned}$$

is p.s.d., so that by Weyl's Monotonicity Theorem (Bhatia (1997)),

$$\mathcal{I}_1(\xi, P) = ch_{\max}\left\{(Q'UQ)^{-1}Q'U^2Q(Q'UQ)^{-1}\right\} \ge ch_{\max}\left\{(Q'_+Q_+)^{-1}\right\}.$$

(ii) We use the following identities, which follow from (10a) and (10b), expressed in the notation preceding the statement of Theorem 3:

$$\mathcal{I}_0(\xi, I_n) = tr\left\{(Q'_+D_+Q_+)^{-1}\right\} \tag{A.5a}$$

$$\mathcal{I}_1(\xi, I_n) = ch_{\max}\left\{(Q'_+D_+Q_+)^{-1}\left(Q'_+D_+^2Q_+\right)(Q'_+D_+Q_+)^{-1}\right\} \tag{A.5b}$$

$$\mathcal{I}_0(\xi, P_0) = tr\left\{(Q'_+Q_+)^{-1}\left(Q'_+D_+^{-1}Q_+\right)(Q'_+Q_+)^{-1}\right\}, \tag{A.5c}$$

$$\mathcal{I}_1(\xi, P_0) = ch_{\max}\left\{(Q'_+Q_+)^{-1}\right\}. \tag{A.5d}$$

To prove (ii) we show that if either inequality fails then one of (14a), (14b) holds – a contradiction. First note that

$$
\begin{aligned}
& \mathcal{I}_0\left(\xi, P_0\right) - \mathcal{I}_0\left(\xi, I_n\right) \\
= \; & tr\left\{\left(Q'_+ Q_+\right)^{-1}\left(Q'_+ D_+^{-1} Q_+\right)\left(Q'_+ Q_+\right)^{-1} - \left(Q'_+ D_+ Q_+\right)^{-1}\right\} \qquad\qquad (\text{A.6}) \\
= \; & tr\left\{\left(Q'_+ Q_+\right)^{-1} Q'_+ D_+^{-1/2}\left[I_q - D_+^{1/2} Q_+\left(Q'_+ D_+ Q_+\right)^{-1} Q'_+ D_+^{1/2}\right] D_+^{-1/2} Q_+\left(Q'_+ Q_+\right)^{-1}\right\},
\end{aligned}
$$

which is non-negative. If the first inequality fails, so that $\mathcal{I}_0\left(\xi, P_0\right) = \mathcal{I}_0\left(\xi, I_n\right)$, then the trace of the p.s.d. matrix at (A.6) is zero, hence all eigenvalues are zero and the matrix is the zero matrix. This is (14a).

That $\mathcal{I}_1\left(\xi, I_n\right) - \mathcal{I}_1\left(\xi, P_0\right) \geq 0$ is the first inequality in (i). If the second inequality of (ii) fails, then $\mathcal{I}_1\left(\xi, I_n\right) = \mathcal{I}_1\left(\xi, P_0\right)$ and their evaluations at (A.5b) and (A.5d) give (14b).

For (iii), that $\Delta_0\left(\varepsilon\right) > 0$ and $\Delta_1\left(\varepsilon\right) > 0$ for sufficiently small ε follow from the continuity of $\mathcal{I}_0\left(\xi, P_\varepsilon\right)$ and $\mathcal{I}_1\left(\xi, P_\varepsilon\right)$ as functions of ε: $\Delta_0\left(\varepsilon\right) \rightarrow \mathcal{I}_0\left(\xi, P_0\right) - \mathcal{I}_0\left(\xi, I_n\right) > 0$ and $\Delta_1\left(\varepsilon\right) = \mathcal{I}_1\left(\xi, I_n\right) - \mathcal{I}_1\left(\xi, P_0\right) > 0$ as $\varepsilon \rightarrow 0$. $\qquad\square$

Proof of Theorem 3. (i) From the first inequality in Lemma 2 (i),

$$
\mathcal{I}_0\left(\xi, I_n\right) = \min_{P>0} \mathcal{I}_0\left(\xi, P\right).
$$

If $P = I_n$ then $U = J'PJ = J'J = D_+ = kI_q$, so that from (A.5b), and the second inequality in Lemma 2(i),

$$
\mathcal{I}_1\left(\xi, I_n\right) = ch_{\max}\left\{\left(Q'_+ Q_+\right)^{-1}\right\} = \min_{P>0} \mathcal{I}_1\left(\xi, P\right).
$$

Now (13) is immediate. If $P_0 = I_n$ then $q = rk\left(P_0\right) = n$, so that all n observations are made at distinct points, hence $D_+ = I_n$ and the design is uniform on its support.

(ii) By Lemma 2(iii) there is $\varepsilon_0 > 0$ for which $\Delta_0\left(\varepsilon\right) > 0$ and $\Delta_1\left(\varepsilon\right) > 0$ when $0 < \varepsilon \leq \varepsilon_0$. For ε in this range,

$$
\mathcal{I}_\nu\left(\xi, I_n\right) - \mathcal{I}_\nu\left(\xi, P_\varepsilon\right) = \nu\left(\Delta_0\left(\varepsilon\right) + \Delta_1\left(\varepsilon\right)\right) - \Delta_0\left(\varepsilon\right) > 0,
$$

for $\nu \in \left(\nu_0, 1\right]$ and $\nu_0 \equiv \Delta_0\left(\varepsilon\right) / \left(\Delta_0\left(\varepsilon\right) + \Delta_1\left(\varepsilon\right)\right)$. $\qquad\square$

Acknowledgements

References

Aitken, A. C. (1935), "On Least Squares and Linear Combinations of Observations," *Proceedings of the Royal Society of Edinburgh*, 55, 42–48.

Bhatia, R. (1997), *Matrix Analysis*, Berlin: Springer.

Cochrane, D. and Orcutt, G. H. (1949), "Application of Least Squares Regression to Relationships Containing Autocorrelated Error Terms," *Journal of the American Statistical Association*, 44, 32–61.

Fang, Z. and Wiens, D. P. (2000), "Integer-Valued, Minimax Robust Designs for Estimation and Extrapolation in Heteroscedastic, Approximately Linear Models," *Journal of the American Statistical Association*, 95, 807–818.

Fomby, T. B., Johnson, S. R., and Hill, R. C. (1984), "Feasible Generalized Least Squares Estimation," in *Advanced Econometric Methods*, New York, NY: Springer.

Heo, G., Schmuland, B., and Wiens, D. P. (2001), "Restricted Minimax Robust Designs for Misspecified Regression Models," *The Canadian Journal of Statistics*, 29, 117–128.

Huber, P. J. (1972), "The 1972 Wald Lecture Robust Statistics: A Review," *The Annals of Mathematical Statistics*, 43, 1041–1067.

Kennedy, J. and Eberhart, R. (1995), "Particle Swarm Optimization," in *Proceedings of ICCN'95 - International Conference on Neural Networks*, WA, Australia, vol. 4: Perth, pp. 1942–1948.

Studden, W. J. (1977), "Optimal Designs for Integrated Variance in Polynomial Regression," in *Statistical Decision Theory and Related Topics II*, eds. Gupta, S. S. and Moore, D., York: Academic Press, pp. 411–420.

Trench, W. F. (1999), "Asymptotic Distribution of the Spectra of a Class of Generalized Kac-Murdock-Szegö Matrices," *Linear Algebra and Its Applications*, 294, 181–192.

Waite, T. W. (2024), "Replication in Random Translation Designs," *Statistics and Probability Letters*, in press, https://doi.org/10.1016/j.spl.2024.110229.

Waite, T. W. and Woods, D. C. (2022), "Minimax Efficient Random Experimental Design Strategies With Application to Model-Robust Design for Prediction," *Journal of the American Statistical Association*, 117, 1452–1465.

Wiens, D. P. (1992), "Minimax Designs for Approximately Linear Regression," *Journal of Statistical Planning and Inference*, 31, 353–371.

— (2000), "Robust Weights and Designs for Biased Regression Models: Least Squares and Generalized M-Estimation," *Journal of Statistical Planning and Inference*, 83, 395–412.

— (2015), "Robustness of Design," *Handbook of Design and Analysis of Experiments*.

— (2018), "I-Robust and D-Robust Designs on a Finite Design Space," *Statistics and Computing*, 28, 241–258.

— (2024a), "Jittering and Clustering: Strategies for the Construction of Robust Designs," *Statistics and Computing*, in press, https://doi.org/10.1007/s11222-024-10436-2.

— (2024b), "A Note on Minimax Robustness of Designs Against Correlated or Heteroscedastic Responses," *Biometrika*, in press, https://doi.org/10.1093/biomet/asae001.