JOURNAL OF STATISTICAL COMPUTATION AND SIMULATION, 2018
VOL. 88, NO. 8, 1437–1453
https://doi.org/10.1080/00949655.2018.1437620

On mitigating collinearity through mixtures

D. R. Jensen[a] and D. E. Ramirez[b]

[a]Department of Statistics, Virginia Tech, Blacksburg, VA, USA; [b]University of Virginia, Charlottesville, VA, USA

ABSTRACT
In linear models having near collinear columns of $\boldsymbol{X}$, ridge and surrogate estimators often are used to mitigate collinearity. A new class of estimators is based on mixtures, either of $\boldsymbol{X}$ and a design minimal in an ordered class or of the Fisher information and a scalar matrix. Comparisons are drawn among choices for the mixing parameter, and the estimators are found to be admissible relative to ordinary least squares. Case studies demonstrate that selected mixture designs are perturbed from the original design to a lesser extent than are those of the surrogate method, while retaining reasonable efficiency characteristics.

ARTICLE HISTORY
Received 15 March 2017
Accepted 3 February 2018

KEYWORDS
Conditioning; ordering by majorization; monotone functions; efficiency indices; design modification

AMS SUBJECT CLASSIFICATIONS
62J05; 62J20

1. Introduction

The models of note are $\{\boldsymbol{Y}_0 = \beta_0 \mathbf{1}_n + \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}\}$ where the columns of $\boldsymbol{X}$, comprising regressors of order $(n \times p)$, have been centred about their means. In addition, elements of $\boldsymbol{Y}_0$ are centred also about their mean $\overline{Y}$ such that

$$\boldsymbol{Y}_0 - \overline{Y}\mathbf{1}_n = \boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} \quad \text{and} \quad \mathbf{1}_n{}'\boldsymbol{X} = \mathbf{0}, \tag{1}$$

with $\mathbf{1}_n{}' = [1, \ldots, 1]$. The assumptions, A1. $\mathrm{E}(\boldsymbol{\epsilon}) = \mathbf{0}$ and $\mathrm{V}(\boldsymbol{\epsilon}) = \sigma^2 \boldsymbol{I}_n$; A2. $\mathcal{L}(\boldsymbol{\epsilon}) = N_n(\mathbf{0}, \sigma^2 \boldsymbol{I}_n)$, are taken to apply, where σ^2 is unity unless specified otherwise. Here $\mathrm{E}(\cdot)$ and $\mathrm{V}(\cdot)$ are the expectation and dispersion operators, and assumption A2 designates the Gaussian law on $\mathbb{R}^n$. The Ordinary Least Squares (OLS) solutions are $\hat{\boldsymbol{\beta}}_L = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y}$. Pervasive issues continue to arise when $\boldsymbol{X}$ is ill-conditioned, i.e. its columns are nearly collinear, resulting in instability of the estimating equations, inflated variances, solutions $\hat{\boldsymbol{\beta}}_L$ having excessive lengths, estimators of doubtful signage and other problematic anomalies.

Hadi [1] in *The International Encyclopedia of Statistical Science* identifies the principal techniques for mitigating collinearity to include the ridge estimators of Hoerl and Kennard [2] and the surrogate estimators of Jensen and Ramirez [3], both essentially data-analytic. Here we focus on the regressors themselves, altered so as to enhance their conditioning, as innate to the structure of collinearity itself. In particular, a recent technique of Jensen and Ramirez [4] is developed further based on mixtures, either of $\boldsymbol{X}$ and an alternative $\boldsymbol{X}_0$ minimal in an ordered class, or a mixture of the ill-conditioned Fisher information matrix $\boldsymbol{X}'\boldsymbol{X}$ with a scalar moment matrix 'ideal' under ordering by Schur majorization.

CONTACT D. E. Ramirez der@virginia.edu University of Virginia, P. O. Box 400137, Charlottesville, VA 22904, USA

Table 1. Given design $\boldsymbol{X} = \boldsymbol{P}\boldsymbol{D}_\xi\boldsymbol{Q}'$ and $\boldsymbol{X}'\boldsymbol{X} = \boldsymbol{Q}\boldsymbol{D}_\lambda\boldsymbol{Q}'$, details follow on their transition into ridge and surrogate regressions, together with the estimators $\{\hat{\boldsymbol{\beta}}_R(k);\ k \geq 0\}$ and $\{\hat{\boldsymbol{\beta}}_S(k);\ k \geq 0\}$.

Item	Ridge regression	Surrogate regression
$\boldsymbol{X}'\boldsymbol{X} \to$	$\boldsymbol{X}'\boldsymbol{X} + k\boldsymbol{I}_p$	$\boldsymbol{X}'\boldsymbol{X} + k\boldsymbol{I}_p$
Special values	$\{\lambda_i \to \lambda_i + k; 1 \leq i \leq p\}$	$\{\xi_i \to (\xi_i^2 + k)^{1/2}; 1 \leq i \leq p\}$
$\boldsymbol{X} \to$	$\boldsymbol{X}$	$\boldsymbol{X}_k = \boldsymbol{P}\,\mathrm{Diag}((\xi_i^2 + k)^{1/2})\boldsymbol{Q}'$
Estimators	$\hat{\boldsymbol{\beta}}_R(k) = (\boldsymbol{X}'\boldsymbol{X} + k\boldsymbol{I}_p)^{-1}\boldsymbol{X}'\boldsymbol{Y}$	$\hat{\boldsymbol{\beta}}_S(k) = (\boldsymbol{X}'\boldsymbol{X} + k\boldsymbol{I}_p)^{-1}\boldsymbol{X}_k{}'\boldsymbol{Y}$

To place this study in perspective, the ill-conditioning of $\boldsymbol{X}'\boldsymbol{X}$ is addressed in ridge regression on perturbing the moment matrix $\{\boldsymbol{X}'\boldsymbol{X} \to \boldsymbol{X}'\boldsymbol{X} + k\boldsymbol{I}_p\}$ with $k \geq 0$, so that all eigenvalues are increased identically, namely $\{\lambda_i \to \lambda_i + k; 1 \leq i \leq p\}$, both the small ones at the root of ill-conditioning as well as large ones not effecting ill-conditioning. Similarly, surrogate regression perturbs the singular values of $\boldsymbol{X} \to \boldsymbol{X}_k$ by the rule $\{\xi_i \to (\xi_i^2 + k)^{1/2}; 1 \leq i \leq p\}$ defined to allow comparisons between the ridge and surrogate procedures, as both have identical moment matrices as seen in Table 1. Moreover, a mixing procedure of Jensen and Ramirez [4] perturbs the eigenvalues of $\boldsymbol{X}'\boldsymbol{X}$ towards a target value, so that the small eigenvalues are increased and the large eigenvalues are decreased. Specifically, $\boldsymbol{X} \to \boldsymbol{X}_t$ is modified such that the eigenvalues of $\boldsymbol{X}'\boldsymbol{X}$ are perturbed by the continuum of rules $\{\lambda_i = \xi_i^2 \to [(1-t)\xi_i^2 + t\bar{\xi}^2]; t \in [0,1]\}$ with target $\bar{\xi}^2$ as the square of the average singular value of $\boldsymbol{X}$. These are called *arithmetic* mixtures owing to the form $\{(1-t)\xi_i^2 + t\bar{\xi}^2; t \in [0,1]\}$.

Against this background, further mixtures are undertaken here: (i) $\boldsymbol{X} \to \boldsymbol{Z}_t$ on perturbing the eigenvalues of $\boldsymbol{X}'\boldsymbol{X}$ by the continuum of rules $\{\lambda_i = \xi_i^2 \to [(1-t)\xi_i^2 + t\bar{\xi}^2; 1 \leq i \leq p\}$ for $t \in [0,1]$ with target as the average eigenvalue; (ii) $\boldsymbol{X} \to \boldsymbol{W}_t$ on perturbing the singular values of $\boldsymbol{X}$ by the rule $\{\xi_i \to [(1-t)\xi_i + t\bar{\xi}]; 1 \leq i \leq p\}$ with target as the average singular value. Details are summarized in Table 3. Theorem 3.2 shows that the design efficiencies $[\mathtt{A}, \mathtt{D}, \mathtt{E}]$ are monotonic in t for $\{\boldsymbol{X} \to \boldsymbol{Z}_t\}$ and $\{\boldsymbol{X} \to \boldsymbol{W}_t\}$ of the present study. In short, altered design points discovered through mixtures may be instructive towards better conditioning in experiments yet to be designed. An outline follows.

Section 2 sets conventions for notation, together with reviews of basic orderings, of ridge and surrogate regressions, of Variance Inflation Factors (*VIF*s), and of the critical Admissibility Criterion for biased alternatives to OLS. Section 3 identifies the mixtures and establishes their admissibility together with other essential properties. Section 4 illustrates the concepts for five data sets known to exhibit collinearity to varying degrees. In comparing ridge, surrogate and three mixtures, all with common *VIF*s, the case studies support two conclusions: (i) one mixture design remains closer to the original design than the surrogate and other mixtures using two metrics to be defined, and (ii) the surrogate design has superior $(\mathtt{A}, \mathtt{D}, \mathtt{E})$ efficiencies in the cases studied. Section 5 draws essential conclusions.

2. Preliminaries

Conventions for notation are followed by surveys of essential supporting topics. These include ridge and surrogate regression, *VIF*s, and elements of Schur majorization.

2.1. Notation

Denote by $\mathbb{R}^p$ the Euclidean p -space; by $\mathbb{R}^p_+$ its positive orthant; by $\mathbb{F}_{n\times p}$ the real $(n \times p)$ matrices of rank $p < n$; by $\mathbb{S}_p$ the real symmetric $(p \times p)$ matrices, with $\mathbb{S}^0_p$, $\mathbb{S}^+_p$ and $\mathbb{D}_p$ as their positive semidefinite, positive definite and diagonal varieties. The transpose, trace and determinant of $\boldsymbol{A}$ are $\boldsymbol{A}'$, $\mathrm{tr}(\boldsymbol{A})$ and $|\boldsymbol{A}|$; and special arrays include the unit vector $\mathbf{1}_p = [1, \ldots, 1]' \in \mathbb{R}^p$, the unit matrix $\boldsymbol{I}_p$ and a typical diagonal matrix $\boldsymbol{D}_\alpha = \mathrm{Diag}(\alpha_1, \ldots, \alpha_p) \in \mathbb{D}_p$. Transformation groups acting on $\mathbb{R}^p$ include the general linear group $\mathcal{G}_p$ and the real orthogonal group $\mathcal{O}_p$. The *spectral decomposition* of $\boldsymbol{A}$ is $\boldsymbol{A} = \sum_{i=1}^p \alpha_i \boldsymbol{q}_i \boldsymbol{q}_i' \in \mathbb{S}^+_p$ with $\lambda(\boldsymbol{A}) = \{\alpha_1 \geq \ldots \geq \alpha_p > 0\}$ as its eigenvalues. The *singular decomposition* of $\boldsymbol{X} \in \mathbb{F}_{n\times p}$ is $\boldsymbol{X} = \sum_{i=1}^p \xi_i \boldsymbol{p}_i \boldsymbol{q}_i' = \boldsymbol{P}\boldsymbol{D}_\xi \boldsymbol{Q}'$ in which $\boldsymbol{P} = [\boldsymbol{p}_1, \ldots, \boldsymbol{p}_p]$ contains the *left singular vectors,* $\boldsymbol{Q} = [\boldsymbol{q}_1, \ldots, \boldsymbol{q}_p] \in \mathcal{O}_p$ contains the *right singular vectors,* and elements of $\boldsymbol{D}_\xi = \mathrm{Diag}(\xi_1, \ldots, \xi_p)$ are its ordered *singular values* under the mapping $\sigma(\boldsymbol{X}) = \{\xi_1 \geq \ldots \geq \xi_p > 0\}$. Denote by $\mathrm{tr}^\dagger(\boldsymbol{X}) = \mathrm{tr}(\boldsymbol{D}_\xi) = \sum_{i=1}^p \xi_i$. Moreover, for subsequent reference let $\mathbb{F}^\tau_{n\times p} = \{\boldsymbol{X} \in \mathbb{F}_{n\times p} : \mathrm{tr}^\dagger(\boldsymbol{X}) = \tau\}$ and $\mathbb{S}^\tau_p = \{\boldsymbol{A} \in \mathbb{S}^+_p : \mathrm{tr}(\boldsymbol{A}) = \tau\}$.

Standard usage refers to independent, identically distributed *(iid)* variates, their cumulative distribution function *(cdf)* and $\mathcal{L}(\boldsymbol{Y})$ as the distribution of $\boldsymbol{Y}$, with $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ as the Gaussian law on $\mathbb{R}^p$ having the mean $\mathrm{E}(\boldsymbol{Y}) = \boldsymbol{\mu}$ and dispersion matrix $\mathrm{V}(\boldsymbol{Y}) = \boldsymbol{\Sigma}$.

Definition 2.1: In regard to the model $\{\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}\}$, the matrix $\boldsymbol{X}$ may be either observational, i.e. concomitant variables observed during the course of an experiment, or as points in the space of the concomitant variables specified by a given design. In either case, $\boldsymbol{X}$ will be called a *design matrix* and $\boldsymbol{X} \to \boldsymbol{X}_\omega$ as its *design modification.*

2.2. Ordered spaces

A partially ordered set $(\mathcal{A}, \succeq_0)$ satisfies the order axioms: (i) antisymmetric, (ii) reflexive and (iii) transitive. It is a *lower semi-lattice* if for elements (x, y) in $\mathcal{A}$, there is a *greatest lower bound* ($glb = x \wedge y$) in $\mathcal{A}$; an *upper semi-lattice* if there is a *least upper bound* ($lub = x \vee y$) in $\mathcal{A}$; and a *lattice* if both a lower and upper semi-lattice. Such spaces are central to this study.

In particular, take the simplex $\mathbb{C}_k = \{\boldsymbol{x} \in \mathbb{R}^k_+ \mid x_1 \geq \ldots \geq x_k\}$ and, for $(\boldsymbol{x}, \boldsymbol{y}) \in \mathbb{C}_k$, suppose that

$$\{x_1 + x_2 + \cdots + x_t \geq y_1 + y_2 + \cdots + y_t;\ 1 \leq t \leq k-1\} \tag{2}$$

$$\{x_1 + x_2 + \cdots + x_k = y_1 + y_2 + \cdots + y_k\}. \tag{3}$$

Then $\boldsymbol{x}$ is said to *majorize* $\boldsymbol{y}$, to be denoted as $\boldsymbol{x} \succeq \boldsymbol{y}$. The functions monotone increasing under $\succcurlyeq$ are called *Schur convex (S -convex)* or *S -concave* if decreasing. Vectors $\boldsymbol{x} \succeq \boldsymbol{y}$ are related as $\boldsymbol{x}\mathbf{P} = \boldsymbol{y}$ through a doubly stochastic matrix $\mathbf{P}$, or the recovery of $\boldsymbol{y}$ from $\boldsymbol{x}$ through a finite number of T-transforms [5].

Recalling $\sigma(\boldsymbol{X}) = (\kappa_1 \geq \ldots \geq \kappa_p)$ and $\lambda(\boldsymbol{A}) = (\alpha_1 \geq \ldots \geq \alpha_p)$ as their respective singular and eigenvalue mappings, we have the following.

Definition 2.2: (i) Let $\mathbb{C}_k^\tau = \{\boldsymbol{x} \in \mathbb{C}_k \mid \sum_{i=1}^k x_i = \tau\}$ together with the ordering $(\mathbb{C}_k^\tau, \succeq)$.

(ii) Let $(\mathbb{F}_{n\times p}^\tau, \succeq_{\mathrm{S}})$ be ordered such that $\boldsymbol{X} \succeq_{\mathrm{S}} \boldsymbol{Y} \in (\mathbb{F}_{n\times p}^\tau, \succeq_{\mathrm{S}})$ if and only if their singular values are ordered by majorization as $\sigma(\boldsymbol{X}) \succeq \sigma(\boldsymbol{Y})$ in $(\mathbb{C}_p^\tau, \succeq)$.

(iii) Let $(\mathbb{S}_p^\tau, \succeq_{\mathrm{S}})$ be ordered such that $\boldsymbol{A} \succeq_{\mathrm{S}} \boldsymbol{B}$ in $(\mathbb{S}_p^\tau, \succeq_{\mathrm{S}})$ if and only if their eigenvalues are ordered by majorization as $\lambda(\boldsymbol{A}) \succeq \lambda(\boldsymbol{B})$ in $(\mathbb{C}_p^\tau, \succeq)$.

(iv) Let $(\mathbb{S}_p^+, \succeq_L)$ be ordered as in [6] such that $\boldsymbol{A} \succeq_L \boldsymbol{B}$ if and only if $(\boldsymbol{A} - \boldsymbol{B}) \in \mathbb{S}_p^0$, with $\boldsymbol{A} \succ_L \boldsymbol{B}$ for $(\boldsymbol{A} - \boldsymbol{B}) \in \mathbb{S}_p^+$.

To continue, essential properties follow, where the *glb* for $(\boldsymbol{X}_\omega, \boldsymbol{X}_\theta)$ is $\boldsymbol{X}_{\mathrm{M}} = \boldsymbol{X}_\omega \wedge \boldsymbol{X}_\theta$ such that both $\boldsymbol{X}_\omega \succeq_{\mathrm{S}} \boldsymbol{X}_{\mathrm{M}}$ and $\boldsymbol{X}_\theta \succeq_{\mathrm{S}} \boldsymbol{X}_{\mathrm{M}}$ in $(\mathbb{F}_{n\times p}, \succeq_{\mathrm{S}})$, and similarly for their *lub*.

Proposition 2.1: (i) *The space* $(\mathbb{C}_k^\tau, \succeq)$ *is a lattice with* $\boldsymbol{x} \wedge \boldsymbol{y}$ *and* $\boldsymbol{x} \vee \boldsymbol{y}$ *as in Equations* (2.1) *and* (2.2) *of Jensen* [7].

(ii) *For* $\boldsymbol{X} = \boldsymbol{P}\,\mathrm{Diag}(\boldsymbol{\omega})\boldsymbol{Q}'$ *and* $\boldsymbol{Z} = \boldsymbol{P}\,\mathrm{Diag}(\boldsymbol{\theta})\boldsymbol{Q}'$ *in* $(\mathbb{F}_{n\times p}^\tau, \succeq_S)$, *then the*

$$glb : \boldsymbol{X} \wedge \boldsymbol{Z} = \boldsymbol{P}\,\mathrm{Diag}(\boldsymbol{\omega} \wedge \boldsymbol{\theta})\boldsymbol{Q}' \quad \text{and} \quad lub : \boldsymbol{X} \vee \boldsymbol{Z} = \boldsymbol{P}\,\mathrm{Diag}(\boldsymbol{\omega} \vee \boldsymbol{\theta})\boldsymbol{Q}' \tag{4}$$

are in $(\mathbb{F}_{n\times p}^\tau, \succeq_S)$.

(iii) *For* $\boldsymbol{A} = \boldsymbol{Q}\,\mathrm{Diag}(\boldsymbol{\alpha})\boldsymbol{Q}'$ *and* $\boldsymbol{B} = \boldsymbol{Q}\,\mathrm{Diag}(\boldsymbol{\beta})\boldsymbol{Q}'$ *in* $(\mathbb{S}_p^\tau, \succeq_S)$, *then* $\boldsymbol{A} \wedge \boldsymbol{B} = \boldsymbol{Q}\,\mathrm{Diag}(\boldsymbol{\alpha} \wedge \boldsymbol{\beta})\boldsymbol{Q}'$ *and* $\boldsymbol{A} \vee \boldsymbol{B} = \boldsymbol{Q}\,\mathrm{Diag}(\boldsymbol{\alpha} \vee \boldsymbol{\beta})\boldsymbol{Q}'$ *are in* $(\mathbb{S}_p^\tau, \succeq_S)$.

(iv) *For* $(\boldsymbol{A}, \boldsymbol{B}) \in (\mathbb{S}_p^+, \succeq_L)$, *then* $\boldsymbol{A} \wedge \boldsymbol{B} = \boldsymbol{Q}\,\mathrm{Diag}(\min\{\alpha_i, \beta_i\}; 1 \le i \le p)\boldsymbol{Q}'$ *and* $\boldsymbol{A} \vee \boldsymbol{B} = \boldsymbol{Q}\,\mathrm{Diag}(\max\{\alpha_i, \beta_i\}; 1 \le i \le p)\boldsymbol{Q}'$.

Proof: Foundations for this work trace to Jensen [7,8]. Conclusion (i) was obtained constructively in 1993 essentially by working backwards in expressions (2) and (3). Conclusion (ii) is from Jensen and Ramirez [4], and Conclusion (iii) from Jensen [9]. Conclusion (iv) was pivotal as established and applied in [10]. ■

Subsequent developments rely heavily on the notion of mixtures. On defining the constant vector $\boldsymbol{c}' = [c, c, \ldots, c] \in \mathbb{R}^p$ as minimal in $(\mathbb{C}_p^\tau, \succeq)$ with $\tau = pc$, the following is basic.

Lemma 2.1: *Take* $\boldsymbol{a}$ *and* $\boldsymbol{c}$ *in* $(\mathbb{C}_p^\tau, \succeq)$ *with* $\tau = \sum_{i=1}^p a_i = pc$, *and consider the 'mixture'* $\boldsymbol{m}(t) = [(1-t)\boldsymbol{a} + t\boldsymbol{c}]$. *Then* $\boldsymbol{m}(t) \succeq \boldsymbol{m}(s)$ *for each* $0 \le t < s \le 1$.

Proof: The differences

$$\begin{aligned}
&[(1-t)a_1 + tc] - [(1-s)a_1 + sc] = (s-t)(a_1 - c) > 0\\
&[(1-t)a_1 + tc + (1-t)a_2 + tc] - [(1-s)a_1 + sc + (1-s)a_2 + sc]\\
&\quad = (s-t)(a_1 + a_2 - 2c) > 0\\
&(\text{by induction: } \{r = 1, 2, \ldots, p\}) = (s-t)(a_1 + \cdots + a_r - rc) > 0
\end{aligned}$$

are essential, where the assertions '> 0' follow since $\boldsymbol{a} \succeq \boldsymbol{c}$ in $(\mathbb{C}_p^\tau, \succeq)$. ■

2.3. Ridge and surrogate regression

For a given parameter $k \geq 0$, both the ridge (RR(k)) and surrogate (SR(k)) estimators modify the ill-conditioned $\boldsymbol{X}'\boldsymbol{X} \to \boldsymbol{X}'\boldsymbol{X} + k\boldsymbol{I}_p$ in the same manner, but solve disparate estimating equations with $\boldsymbol{X} \to \boldsymbol{X}_k$ in surrogate regression. Details are supplied in Table 1. In considering collinearity among the columns of $\boldsymbol{X}$, we may assume that not all singular values are equal and, as before, the singular values $\sigma(\boldsymbol{X}) = \{\xi_1 \geq \ldots \geq \xi_p\}$ are arrayed as $\boldsymbol{D}_\xi = \mathrm{Diag}(\xi_1, \ldots, \xi_p)$, and the eigenvalues as $\lambda(\boldsymbol{X}'\boldsymbol{X}) = \boldsymbol{D}_\lambda = \mathrm{Diag}(\lambda_1, \ldots, \lambda_p)$ with $\{\lambda_i = \xi_i{}^2; 1 \leq i \leq p\}$.

As the ridge and surrogate solutions are not equivariant under scaling, it is conventional to scale $\boldsymbol{V} = (\boldsymbol{X}'\boldsymbol{X})^{-1}$ into its *correlation form* by centering, then scaling, each column of $\boldsymbol{X}$ to unit lengths. Then the variances $\mathrm{Var}(\hat{\beta}_{Lj})$ are equal, which in turn justifies the single perturbation parameter $k > 0$ for all variables in $\boldsymbol{X}'\boldsymbol{X} \to \boldsymbol{X}'\boldsymbol{X} + k\boldsymbol{I}_p$. In case studies to follow, we adhere to these conventions.

2.4. Variance inflation factors

The *VIF*s serve to gauge effects of ill-conditioning on variances of the estimators. The OLS solutions $\hat{\boldsymbol{\beta}}_L{}' = [\,\hat{\beta}_{L1}, \ldots, \widehat{\beta}_{Lp}]$ with $\mathrm{V}(\hat{\boldsymbol{\beta}}_L) = \boldsymbol{V}$ have $\{\mathrm{Var}(\hat{\beta}_{Lj}) = v_{jj}; 1 \leq j \leq p\}$ as actual values. Were the columns of $\boldsymbol{X}$ to be orthogonal, thus 'ideal', then $\{\mathrm{Var}(\hat{\beta}_{Lj}) = w_{jj}{}^{-1}; 1 \leq i \leq p\}$ with $\boldsymbol{X}'\boldsymbol{X} = \mathrm{Diag}(w_{11}, \ldots, w_{pp})$. Accordingly, *VIF*s are defined as $\{VIF(\hat{\beta}_{Lj}) = v_{jj}w_{jj}; 1 \leq j \leq p\}$, namely, ratios of actual to 'ideal' variances. Marquardt and Snee [11] have identified *VIF* as 'the best single measure of the conditioning of the data'.

To achieve the intended *improvement* in conditioning, one would expect that $VIF(\hat{\beta}_j(k)) \to 1$ as $k \to \infty$ for ridge and surrogate solutions. Extensions for ridge and surrogate *VIF*s are given in [12], where it is shown, contrary to expectations, that $VIF(\widehat{\beta}_{Rj}(k)) \to VIF(\hat{\beta}_{Lj})$ as $k \to \infty$. Thus for large k the ridge *VIF*s return to the original ill-conditioned OLS values. On the other hand, it was shown that $VIF(\hat{\beta}_{Sj}(k)) \to 1$ monotonically as $k \to \infty$, so that the surrogate *VIF*s eventually converge to their minimal values. Moreover, *VIF*s are unambiguous for models without intercept for reasons given in [13]. Accordingly, we continue to focus on the model $\{\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}\}$ of Equation (1).

It remains to ask, how large are *VIF*s to be of consequence? Current rules-of-thumb are those exceeding 10 or even 4 [14–16]. A recurring problem centres on the choices for k. In the case studies to be reported, we choose k so as to achieve $\{\max\{VIF(\hat{\beta}_j(k)) : 1 \leq j \leq p\} = 10\}$. Fixing to a constant serves to equilibrate diverse models for comparative purposes. A choice alternative to 10 is examined in Section 5.4.

2.5. Admissibility criterion

Biased estimators are gauged via their Mean Squared Errors, $\mathtt{MSE} = \mathrm{Variance} + \mathrm{Bias}^2$. Hoerl and Kennard [2] established ridge estimators to be MSE *-admissible*, i.e. $\mathtt{MSE}(\hat{\boldsymbol{\beta}}_R(k)) < \mathtt{MSE}(\hat{\boldsymbol{\beta}}_L)$ for some $k \in (0, \infty)$, assuring a reduction in MSE from that of OLS. This result is shown for surrogate regression in [12]. As a caution, however, those authors reported in [17] the existence of cross-over values k_0 for which, if $k > k_0$ then $\mathtt{MSE}(\hat{\boldsymbol{\beta}}_R(k)) > \mathtt{MSE}(\hat{\boldsymbol{\beta}}_L)$, so that all $\{\hat{\boldsymbol{\beta}}_R(k);\ k > k_0\}$ are inadmissible. The stronger

result, that $\mathtt{MSE}(\hat{\boldsymbol{\beta}}_S(k)) \leq \mathtt{MSE}(\hat{\boldsymbol{\beta}}_R(k))$ for all $k \geq 0$ from that study, effectively supplants ridge estimators for prediction. Mixture estimators to follow will be shown to satisfy the *Admissibility Criterion,* namely $(\mathrm{d}/\mathrm{d}k)\mathtt{MSE}(k)|_{k=0} < 0$.

3. Estimation via mixtures

3.1. Overview

Our developments support the use of designs $\boldsymbol{X} \in (\mathbb{F}^{\tau}_{n\times p}, \succeq_{\mathrm{S}})$ whose singular values are 'smoother' than other designs in the sense of majorization. A transformation $\boldsymbol{X} \to \boldsymbol{X}_\omega$ is sought to enhance the conditioning of $\boldsymbol{X}$, as might its Fisher information matrix $\boldsymbol{F}_I(\boldsymbol{X}_\omega) = \boldsymbol{X}_\omega{}'\boldsymbol{X}_\omega$. Initiated briefly in [4], this approach is extended here to include alternative choices for mixtures; to an examination of their capacities to modulate ill-conditioning, and thus adverse effects in linear inference; and to a comparison of these *inter se* and with ridge and surrogate methods. Consider the following.

Definition 3.1: The collection $\{\boldsymbol{L}_i = \boldsymbol{p}_i\boldsymbol{q}_i{}'; 1 \leq i \leq p\}$ may be viewed as *frames* of order $(n \times p)$, i.e. basis elements in $\mathbb{F}_{n\times p}$ supporting the design $\boldsymbol{X} = \sum_{i=1}^{p} \xi_i \boldsymbol{L}_i$.

Then taking $\boldsymbol{X} \to \boldsymbol{X}_\omega = \boldsymbol{P}\mathrm{D}_\omega\boldsymbol{Q}'$ as in Table 2 amounts to representing $\boldsymbol{X}_\omega$ while retaining this basis. To proceed, we construct altered designs, together with the linear estimator $\hat{\boldsymbol{\beta}}_\omega$, its expectation, its dispersion matrix, bias and $\mathtt{MSE}$, on recalling that biased estimators typically are assessed by their $\mathtt{MSE}$ at $\mathrm{E}(\hat{\boldsymbol{\beta}}_\omega) = \boldsymbol{\beta}_0$, namely $\mathtt{MSE}(\hat{\boldsymbol{\beta}}_\omega) = \mathrm{tr}(\boldsymbol{\Sigma}_\omega) + (\boldsymbol{\beta} - \boldsymbol{\beta}_0)'(\boldsymbol{\beta} - \boldsymbol{\beta}_0)$. The ordered eigenvalues of $\mathrm{V}(\hat{\boldsymbol{\beta}}_\omega) = (\boldsymbol{X}_\omega{}'\boldsymbol{X}_\omega)^{-1}$ are $\lambda(\boldsymbol{\Sigma}_\omega) = (1/\omega_p^2 \geq \ldots \geq 1/\omega_1^2)$.

3.2. The mixtures

Our choices are summarized in Table 3, the first from Section 4.4 of Jensen and Ramirez [4]. It is seen from $\{\boldsymbol{F}_I(\boldsymbol{X}_\omega) = [(1-t)\boldsymbol{X}'\boldsymbol{X} + t\,(\kappa\boldsymbol{I}_p)]; t \in [0,1]\}$ that $\boldsymbol{F}_I(\boldsymbol{X}_\omega)$ is drawn towards the perfectly conditioned $\kappa\boldsymbol{I}_p$, and that $\boldsymbol{X} \to \boldsymbol{X}_\omega|_{t=1}$. In the case studies to be reported, t is chosen so as to bound *VIF*s within accepted ranges. All solutions offer a continuum of type $\{\boldsymbol{X}_t;\ t \in [0,1]\}$ as prospects for improved conditioning. Note that the successive scalar values in the displays are $\{t\bar{\xi}^2, t\bar{\xi^2}, t\bar{\xi}\}$ with $\bar{\xi}$ as the average, $\bar{\xi}^2$ as its square and $\bar{\xi^2}$ as the average of the squared singular values of $\boldsymbol{X}$, such that $\{\bar{\xi}^2 < \bar{\xi^2}\}$. Note that the eigenvalues of $\boldsymbol{Z}_t{}'\boldsymbol{Z}_t$, and the singular values of $\boldsymbol{W}_t$, sum to $p\bar{\xi^2}$ and $p\bar{\xi}$, respectively.

Table 2. The modified $\boldsymbol{X} = \boldsymbol{P}\boldsymbol{D}_\xi\boldsymbol{Q}' \to \boldsymbol{X}_\omega = \boldsymbol{P}\mathrm{D}_\omega\boldsymbol{Q}'$, together with the resulting linear estimators and their essential properties.

Modified regressions	Properties
$\boldsymbol{X} \to \boldsymbol{X}_\omega = \boldsymbol{P}\mathrm{D}_\omega\boldsymbol{Q}'$	$\mathrm{E}(\hat{\boldsymbol{\beta}}_\omega) = \boldsymbol{Q}\mathrm{D}_\omega^{-1}\boldsymbol{D}_\xi\boldsymbol{Q}'\boldsymbol{\beta}$
$\hat{\boldsymbol{\beta}}_\omega = \boldsymbol{Q}\mathrm{D}_\omega^{-1}\boldsymbol{P}'\boldsymbol{Y}$	$\mathrm{V}(\hat{\boldsymbol{\beta}}_\omega) = \boldsymbol{Q}\mathrm{D}_\omega^{-2}\boldsymbol{Q}' = \boldsymbol{\Sigma}_\omega$
$\mathtt{MSE} = \sum_{i=1}^{p}[\frac{1}{\omega_i^2} + \theta_i^2(\frac{\xi_i - \omega_i}{\omega_i})^2]; \boldsymbol{\theta} = \boldsymbol{Q}'\boldsymbol{\beta}$	

Table 3. Mixtures generated by a design $\boldsymbol{X} = \boldsymbol{P}\boldsymbol{D}_\xi\boldsymbol{Q}'$ and its $\boldsymbol{F}_I(\boldsymbol{X}) = \boldsymbol{Q}\boldsymbol{D}_\xi{}^2\boldsymbol{Q}'$, together with their altered design, $\boldsymbol{F}_I(\cdot)$ matrix and admissibility criterion.

Item	Design	Fisher Information Matrix	$\frac{\mathrm{d}}{\mathrm{d}t}\mathtt{MSE}(\hat{\boldsymbol{\beta}}_\omega(t))\vert_{t=0}$
$\boldsymbol{X}_t$	$\boldsymbol{P}\,\mathrm{Diag}([(1-t)\xi_i^2 + t\bar{\xi}^2]^{1/2})\boldsymbol{Q}'$	$\boldsymbol{Q}\,\mathrm{Diag}([(1-t)\xi_i^2 + t\bar{\xi}^2])\boldsymbol{Q}'$	$\sum_{i=1}^p \frac{(\xi_i^2-\bar{\xi}^2)}{\xi_i^4}$
$\boldsymbol{Z}_t$	$\boldsymbol{P}\,\mathrm{Diag}([(1-t)\xi_i^2 + t\bar{\xi}^2]^{1/2}\boldsymbol{Q}'$	$\boldsymbol{Q}\,\mathrm{Diag}([(1-t)\xi_i^2 + t\bar{\xi}^2])\boldsymbol{Q}'$	$\sum_{i=1}^p \frac{(\xi_i^2-\bar{\xi}^2)}{\xi_i^4}$
$\boldsymbol{W}_t$	$\boldsymbol{P}\,\mathrm{Diag}([(1-t)\xi_i + t\bar{\xi}])\boldsymbol{Q}'$	$\boldsymbol{Q}\,\mathrm{Diag}([(1-t)\xi_i + t\bar{\xi}]^2)\boldsymbol{Q}'$	$2\sum_{i=1}^p \frac{(\xi_i-\bar{\xi})}{\xi_i^3}$

3.3. Admissible solutions

Before examining their properties further, it is essential first to determine whether $\{\boldsymbol{X}_t, \boldsymbol{Z}_t, \boldsymbol{W}_t\}$ are $\mathtt{MSE}$-admissible and thus are viable alternatives to $\mathtt{OLS}$. We have the following.

Theorem 3.1: *Each of the family of designs $\{\boldsymbol{X}_t, \boldsymbol{Z}_t, \boldsymbol{W}_t;\ t \in [0,1]\}$ has $\mathtt{MSE}(\hat{\boldsymbol{\beta}}_\omega(t))$ to be decreasing at $t = 0$ and so satisfies the Admissibility Criterion to give $\mathtt{MSE}(\hat{\boldsymbol{\beta}}_\omega(t)) < \mathtt{MSE}(\hat{\boldsymbol{\beta}}_L)$ for some $t \in [0,1]$.*

Proof: Designs in the collection $\{(\boldsymbol{Z}_t, \boldsymbol{W}_t);\ t \in [0,1]\}$ were constructed such that $\mathrm{tr}(\boldsymbol{Z}_t{}'\boldsymbol{Z}_t) = p\bar{\xi}^2$ and $\mathrm{tr}^\dagger(\boldsymbol{X}_t) = p\bar{\xi}$. Chebyshev's sum inequality applies directly to show

$$\frac{\sum_{i=1}^p \frac{1}{\xi_i^2}}{\sum_{i=1}^p \frac{1}{\xi_i^4}} \le \bar{\xi}^2 \quad \text{and} \quad \frac{\sum_{i=1}^p \frac{1}{\xi_i^2}}{\sum_{i=1}^p \frac{1}{\xi_i^3}} \le \bar{\xi} \tag{5}$$

implying for $\{\boldsymbol{Z}_t, \boldsymbol{W}_t\}$ that $(\mathrm{d}/\mathrm{d}t)\mathtt{MSE}(\hat{\boldsymbol{\beta}}_\omega(t))|_{t=0} < 0$. For $\{\boldsymbol{X}_t;\ t \in [0,1]\}$, the Method of Lagrange Multipliers as in Chapter 19 of Cvetkovski [18] gives $\max\{(\sum_{i=1}^p (1/\xi_i^2))/(\sum_{i=1}^p (1/\xi_i^4))\} = \bar{\xi}^2$ which is the required bound in order that $(\mathrm{d}/\mathrm{d}t)\mathtt{MSE}(\hat{\boldsymbol{\beta}}_\omega(t))|_{t=0} < 0$ for $\boldsymbol{X}_t$, to conclude our proof. ■

3.4. Properties of mixture estimators

Standard criteria for evaluating the design $\boldsymbol{X}_\omega$ include the $\{\mathtt{A}, \mathtt{D}, \mathtt{E}\}$ efficiency indices, where $\mathtt{A} = \mathrm{tr}(\boldsymbol{\Sigma}_\omega)$; $\mathtt{D} = \log(|\boldsymbol{\Sigma}_\omega|)$ and $\mathtt{E} = \lambda_1(\boldsymbol{\Sigma}_\omega)$, together with the condition number $\kappa(\boldsymbol{X}_\omega) = \lambda_1/\lambda_p$ and Mauchly's [19] *Sphericity Criterion*

$$\mathtt{M}(\boldsymbol{X}_\omega) = p^p \prod_{i=1}^p \left(\frac{1}{\omega_i^2}\right) \Big/ \left[\sum_{i=1}^p \left(\frac{1}{\omega_i^2}\right)\right]^p$$

as a function of the ratio of the geometric mean to the arithmetic mean of the eigenvalues of $\boldsymbol{\Sigma}_\omega$. Here with $\mathtt{A} = \mathrm{tr}(\boldsymbol{\Sigma}_\omega) = \sum \mathrm{Var}(\widehat{\beta}_i)$, small values of $\{\kappa, \mathtt{A}, \mathtt{D}, \mathtt{E}\}$ reflect well-conditioned models. On the other hand, $\mathtt{M}(\boldsymbol{X}_\omega)$ serves to gauge the non-sphericity of contours of the Gaussian density of $\hat{\boldsymbol{\beta}}_\omega$, taking the value $\mathtt{M} = 1.0$ when well conditioned and spherical, and $\mathtt{M} < 1.0$ otherwise.

To continue, we examine properties of $\{(\hat{\boldsymbol{\beta}}(\boldsymbol{W}_t), \hat{\boldsymbol{\beta}}(\boldsymbol{X}_t), \hat{\boldsymbol{\beta}}(\boldsymbol{Z}_t));\ t \in [0,1]\}$ together with their dispersion matrices $\{\boldsymbol{\Sigma}_W(t), \boldsymbol{\Sigma}_X(t), \boldsymbol{\Sigma}_Z(t);\ t \in [0,1]\}$. Some properties may be

viewed as mappings $\mathbb{F}^{\tau}_{n\times p} \to \mathbb{R}^1_+$; others as $\mathbb{S}^{\tau}_p \to \mathbb{R}^1_+$. The dichotomy rests on differing structural features in individual cases. The following theorem asserts for $\boldsymbol{W}_t$ and $\boldsymbol{Z}_t$ that the $\{\kappa, \mathtt{A}, \mathtt{D}, \mathtt{E}\}$ criteria are monotone decreasing, and $\mathtt{M}(\cdot)$ increasing, for $t \uparrow\in [0,1]$.

Theorem 3.2: *Consider the ensembles $\{(\boldsymbol{W}_t, \boldsymbol{Z}_t);\ t \in [0,1]\}$, and their inverse Fisher information matrices as $\mathrm{V}(\hat{\boldsymbol{\beta}}_W(t)) = \boldsymbol{\Sigma}_W(t)$ and $\mathrm{V}(\hat{\boldsymbol{\beta}}_Z(t)) = \boldsymbol{\Sigma}_Z(t)$, together with invariant functions taking $\mathbb{F}^{\tau}_{n\times p} \to \mathbb{R}^1_+$ or $\mathbb{S}^{\tau}_p \to \mathbb{R}^1_+$. The condition numbers $\kappa(\boldsymbol{W}_t)$, the operators $\Delta(\boldsymbol{W}_t) = |\boldsymbol{\Sigma}_W(t)|$, $\Gamma(\boldsymbol{W}_t) = tr(\boldsymbol{\Sigma}_W(t))$ and the eigenvalues $\lambda(\boldsymbol{\Sigma}_W(t)) = [\lambda_1 \geq \ldots \geq \lambda_k]$.*

(i) *Then $\{\kappa(\boldsymbol{W}_t), \Delta(\boldsymbol{W}_t), \Gamma(\boldsymbol{W}_t), \lambda_1(\boldsymbol{\Sigma}_W(t))\}$, as mappings $\mathbb{F}^{\tau}_{n\times p} \to \mathbb{R}^1_+$, all decrease monotonically in $\mathbb{R}^1_+$ for $t \uparrow\in [0,1]$.*

(ii) *The $(\kappa, \mathtt{A}, \mathtt{D}, \mathtt{E})$-efficiency indices for $\hat{\boldsymbol{\beta}}(\boldsymbol{W}_t)$, considered as functions on $(\mathbb{F}^{\tau}_{n\times p}, \succeq)$, all decrease monotonically in $\mathbb{R}^1_+$ for $t \uparrow\in [0,1]$, whereas $\lambda_p(\boldsymbol{\Sigma}_W(t))$ increase monotonically in $\mathbb{R}^1_+$ for $t \uparrow\in [0,1]$.*

(iii) *Mauchly's criteria $\mathtt{M}(\boldsymbol{W}_t)$ for $p = 2$, as mappings $\mathbb{F}^{\tau}_{n\times p} \to \mathbb{R}^1_+$, increase monotonically in $\mathbb{R}^1_+$ for $t \uparrow\in [0,1]$, i.e. for $s > t$, the Gaussian contours of $\hat{\boldsymbol{\beta}}(\boldsymbol{W}_t)$ are less spherical than those of $\hat{\boldsymbol{\beta}}(\boldsymbol{W}_s)$.*

(iv) *For $\hat{\boldsymbol{\beta}}_Z(t)$, the $(\kappa, \mathtt{A}, \mathtt{D}, \mathtt{E})$-efficiency indices, when considered as functions $\mathbb{S}^{\tau}_p \to \mathbb{R}^1_+$, all decrease monotonically in $\mathbb{R}^1_+$ for $t \uparrow\in [0,1]$.*

(v) *Mauchly's $\mathtt{M}(\boldsymbol{Z}_t)$, here mapping $\mathbb{S}^{\tau}_p \to \mathbb{R}^1_+$, increase monotonically in $\mathbb{R}^1_+$ as $t \uparrow\in [0,1]$, i.e. for $s > t$, the Gaussian contours of $\hat{\boldsymbol{\beta}}(\boldsymbol{Z}_t)$ are less spherical than those of $\hat{\boldsymbol{\beta}}(\boldsymbol{Z}_s)$.*

(vi) *The minimal eigenvalues $\lambda_p(\boldsymbol{\Sigma}_Z(t))$ as mappings $\mathbb{S}^{\tau}_p \to \mathbb{R}^1_+$ increase monotonically in $\mathbb{R}^1_+$ for $t \uparrow\in [0,1]$.*

Proof: In regard to $\boldsymbol{W}_t$, the collection $\{\boldsymbol{W}_t;\ t \in [0,1]\}$ is constructed with $\mathrm{tr}^{\dagger}(\boldsymbol{W}_t) = \sum \xi_i = p\bar{\xi}$ so Lemma 2.1 can be applied to establish that singular values satisfy $\sigma(\boldsymbol{W}_t) \succeq \sigma(\boldsymbol{W}_s)$ in $(\mathbb{C}^{\tau}_p, \succeq)$ for $\{0 \leq t < s \leq 1\}$. Then $\boldsymbol{W}_t \succeq \boldsymbol{W}_s$ in $(\mathbb{F}^{\tau}_{n\times p}, \succeq)$, and properties (i)–(iii) given for $\hat{\boldsymbol{\beta}}_W(t)$ follow from Theorem 4.2 of Jensen and Ramirez [4]. To continue, the ensemble $\{\boldsymbol{Z}_t;\ t \in [0,1]\}$ was constructed so that $\mathrm{tr}(\boldsymbol{Z}_t'\boldsymbol{Z}_t) = \sum \xi_i^2 = p\bar{\xi}^2$, so Lemma 2.1 again may be applied on taking $\boldsymbol{m}(t) = [(1-t)\boldsymbol{a} + t\boldsymbol{c}]$ with $\boldsymbol{a} = \mathrm{Diag}(\xi_1^2, \ldots, \xi_p^2)$ and $\boldsymbol{c} = \mathrm{Diag}(\bar{\xi}^2, \ldots, \bar{\xi}^2)$, so that $\boldsymbol{m}(t) \succeq \boldsymbol{m}(s)$ in $(\mathbb{S}^{\tau}_p, \succeq)$ as in Definition 2.2(ii), that is

$$\mathrm{Diag}([(1-t)\xi_i^2 + t\bar{\xi}^2]) \succeq \mathrm{Diag}([(1-s)\xi_i^2 + s\bar{\xi}^2]),\ \text{equivalently, } \boldsymbol{Z}_t'\boldsymbol{Z}_t \succeq \boldsymbol{Z}_s'\boldsymbol{Z}_s \in (\mathbb{S}^{\tau}_p, \succeq) \tag{6}$$

for each $\{0 \leq t < s \leq 1\}$. Theorem 1 of Jensen [9] asserts that the criteria $(\kappa, \mathtt{A}, \mathtt{D}, \mathtt{E})$ for design $\boldsymbol{Z}_t$ are all monotone decreasing as $t \uparrow\in [0,1]$, as asserted in conclusions (iv) and (vi), while Mauchly's $\mathtt{M}(\cdot)$ in conclusion (v) increases monotonically to 1. ■

Remark 3.1: It is essential to note that $\boldsymbol{X}_t$ thus far has been omitted, as it is not amenable to the foregoing analyses. Specifically, neither the sums of singular values of $\boldsymbol{X}_t$ nor the sums of eigenvalues of $\boldsymbol{X}_t'\boldsymbol{X}_t$ are constant as t ranges over $[0,1]$. In short, $\boldsymbol{X}_s$ and $\boldsymbol{X}_t$ will not be comparable when they lie in different spaces $(\mathbb{F}^{\tau}_{n\times p}, \succeq)$ and $(\mathbb{F}^{\tau\dagger}_{n\times p}, \succeq)$. This may explain

anomalies found in the numerical studies to follow. Nonetheless, common ground is found in the following.

Theorem 3.3: *Consider the ensembles* $\{(\boldsymbol{W}_t, \boldsymbol{X}_t, \boldsymbol{Z}_t);\ t \in [0,1]\}$, *and their Fisher information matrices* $\{\boldsymbol{W}_t'\boldsymbol{W}_t, \boldsymbol{X}_t'\boldsymbol{X}_t, \boldsymbol{Z}_t'\boldsymbol{Z}_t\}$ *and inverses* $\{\boldsymbol{\Sigma}_W, \boldsymbol{\Sigma}_X, \boldsymbol{\Sigma}_Z\}$, *together with the Loewner* [6] *ordering* $(\mathbb{S}_p^+, \succeq_L)$ *as in Definition* 2.2(iii).

(i) *Then* $\{\boldsymbol{W}_t'\boldsymbol{W}_t \preceq_L \boldsymbol{X}_t'\boldsymbol{X}_t \preceq_L \boldsymbol{Z}_t'\boldsymbol{Z}_t$ *in* $(\mathbb{S}_p^+, \succeq_L)\}$.
(ii) *Equivalently* $\{\boldsymbol{\Sigma}_W \succeq_L \boldsymbol{\Sigma}_X \succeq_L \boldsymbol{\Sigma}_Z$ *in* $(\mathbb{S}_p^+, \succeq_L)\}$.
(iii) *Each efficiency index* $(\kappa, \mathtt{A}, \mathtt{D}, \mathtt{E})$ *in* $\mathbb{R}_+^1$ *is ordered as are* $\{\boldsymbol{\Sigma}_W \succeq_L \boldsymbol{\Sigma}_X \succeq_L \boldsymbol{\Sigma}_Z$ *in* $(\mathbb{S}_p^+, \succeq_L)\}$.

Proof: Recall that $\boldsymbol{A} \succeq_L \boldsymbol{B}$ in $(\mathbb{S}_p^+, \succeq_L)$ if and only if their eigenvalues are pairwise ordered as $\{\lambda_i(\boldsymbol{A}) \geq \lambda_i(\boldsymbol{B}); 1 \leq i \leq p\}$. A direct calculation establishes the Mixture Inequality, namely $\{(1-t)a + tb \leq \sqrt{(1-t)a^2 + tb^2};\ a > 0,\ b > 0,\ t \in [0,1]\}$. Thus each ordered singular value of $\boldsymbol{W}_t$ is dominated by the corresponding singular value of $\boldsymbol{X}_t$ and thus $\{\boldsymbol{W}_t'\boldsymbol{W}_t \preceq_L \boldsymbol{X}_t'\boldsymbol{X}_t$ in $(\mathbb{S}_p^+, \succeq_L)\}$. Moreover, from Table 3 the difference

$$[\lambda(\boldsymbol{X}_t'\boldsymbol{X}_t) - \lambda(\boldsymbol{Z}_t'\boldsymbol{Z}_t)] = [(1-t)\xi_i^2 + t\bar{\xi}^2] - [(1-t)\xi_i^2 + t\bar{\xi}^2] = t(\bar{\xi}^2 - \bar{\xi}^2) < 0 \quad (7)$$

holds for each $\{i = 1, 2, \ldots, p\}$, so that $\boldsymbol{X}_t'\boldsymbol{X}_t \preceq_L \boldsymbol{Z}_t'\boldsymbol{Z}_t$, to give conclusion (i) and the equivalent conclusion (ii). Conclusion (iii) follows as an immediate consequence. ■

4. Case studies

4.1. Basics

Data from the literature identified as ill-conditioned are re-examined next, again taking the regressors to be centred and scaled. The form $\{\boldsymbol{X}_k'\boldsymbol{X}_k = \boldsymbol{X}'\boldsymbol{X} + k\boldsymbol{I}_p\}$ is identical for ridge and surrogate models, having the same *VIF*s but without an underlying ridge design. Accordingly, the surrogate $\boldsymbol{X}_k$, but not ridge, is to be compared with the mixture designs $\{\boldsymbol{X}_t, \boldsymbol{Z}_t, \boldsymbol{W}_t;\ t \in [0,1]\}$. Regarding choices for $\{k \in [0,\infty);\ t \in [0,1]\}$, these are determined so that $\max\{VIF_i : 1 \leq i \leq p\} = 10$ in order to standardize consistently across a diversity of ill-conditioned cases. The software package Maple supports all computations. Observe that $\{\boldsymbol{X}_k, \boldsymbol{X}_t, \boldsymbol{Z}_t, \boldsymbol{W}_t\}$ all retain the basic frames of Definition 3.1 for spanning $\mathbb{F}_{n\times p}$, taking these as signature to each design space itself. Moreover, the proximity of $\boldsymbol{X}$ to $\boldsymbol{X}_\omega$ under $\{\boldsymbol{X} \to \boldsymbol{X}_\omega\}$ is the subject of the following.

Remark 4.1: (i) Venues for near collinearity may arise through constraints among the regressors, thus precluding as infeasible some combinations of points in the space of regressors.

(ii) If indeed $\boldsymbol{X}$ is feasible though ill-conditioned, then seeking a nearby $\{\boldsymbol{X} \to \boldsymbol{X}_\omega\}$ holds promise for a feasible version with enhanced conditioning.

In studies to follow design characteristics are listed, to include $VF_M = \max\{VIF; 1 \leq i \leq p\}$, the efficiency indices $\{\kappa, \mathtt{A}, \mathtt{D}, \mathtt{E}\}$ and Mauchly's $\mathtt{M}(\cdot)$. In addition, in keeping

with Remark 4.1, the displacement of $\boldsymbol{X} = [x_{ij}]$ to its modified version $\boldsymbol{X}_\omega = [\,x_{ij}(\omega)]$ is gauged by the *mean absolute deviation,* namely $\mathtt{MAD}(\boldsymbol{X}_\omega) = (1/np)\sum|x_{ij}(\omega) - x_{ij}|$, and by $\Delta(\boldsymbol{X}_\omega) = \sum|\xi_j(\omega) - \xi_j|$ as the discrepancy between their singular values. The latter is invariant under left and right unitary operators, and thus is independent of the basis elements for $\mathbb{F}_{n\times p}$ as in Definition 3.1. Further criteria include the correlations $\rho(\boldsymbol{Y}, \widehat{\boldsymbol{Y}}_\omega)$ with $\widehat{\boldsymbol{Y}}_\omega = \boldsymbol{X}_\omega\hat{\boldsymbol{\beta}}_\omega$, larger correlations reflecting greater integrity in predicting $\boldsymbol{Y}$ through $\hat{\boldsymbol{\beta}}_\omega$.

We refrain from a detailed evaluation of each case *in situ.* Instead, a comprehensive comparison across cases seems more informative as given in the Conclusions. In keeping with the foregoing issues, for each case study we list the numerical diagnostics as follow:

$$[VF_M,\ \kappa,\ \mathtt{A},\ \mathtt{D},\ \mathtt{E},\ \mathtt{M},\ \rho,\ \mathtt{MAD},\ \Delta], \tag{8}$$

where $\mathtt{D} = \log|\boldsymbol{\Sigma}_\omega|$. Recall that the condition number κ and the efficiency indices $(\mathtt{A}, \mathtt{D}, \mathtt{E})$ ideally would be small, whereas the ellipticity index $\mathtt{M}(\cdot)$ would increase towards unity and circular contours in well-conditioned cases.

4.2. Acetylene Data: Marquardt and Snee [11]

For the Five-Coefficient Reduced Quadratic Model with $n = 16$ and $p = 5$, the explanatory variables are: x_1 reactor temperature; x_2 ratio of H_2 to n -heptone; x_3 contact time; x_1x_2 interaction; x_1^2 squared temperature and with y the conversion percentage of n-heptone to acetylene. Table 4 reports values for the original $\boldsymbol{X}$, the surrogate $\boldsymbol{X}_k$ and designs $\{\boldsymbol{X}_t, \boldsymbol{Z}_t, \boldsymbol{W}_t;\ t \in [\,0, 1]\}$. Values include the perturbation parameters (either k or t) and other quantities listed in expression (8).

The parameters have been computed with $VF_M = 10$ to allow for comparisons. All designs show the expected improvement in condition number and design criteria. We observe that $\boldsymbol{Z}_t$ has perturbed the original design $\boldsymbol{X}$ the least with $\mathtt{MAD}(\boldsymbol{Z}_t) = 0.0287$ and $\Delta(\boldsymbol{Z}_t) = 0.6014$. On the other hand, the surrogate design $\boldsymbol{X}_k$ has superior $\{\mathtt{A}, \mathtt{D}, \mathtt{E}\}$ values. Using the Acetylene Data, Table 5 shows that the $(\mathtt{A}, \mathtt{D})$-efficiency indices, when viewed as functions $\mathbb{S}_p^+ \to \mathbb{R}_+^1$, decrease monotonically in $\mathbb{R}_+^1$ for $t \uparrow\in [0, 1]$ for the ensembles $\{\boldsymbol{Z}_t, \boldsymbol{W}_t;\ t \in [0, 1]\}$ as reported in Theorem 3.2. However, the family $\{\boldsymbol{X}_t;\ t \in [0, 1]\}$ lacks monotonicity for the $(\mathtt{A}, \mathtt{D})$-efficiency indices as t increases in $[0, 1]$. Theorem 3.3 asserts that the design efficiencies will be ordered as are $\{\boldsymbol{\Sigma}_W \succeq_L \boldsymbol{\Sigma}_X \succeq_L \boldsymbol{\Sigma}_Z\}$ as demonstrated in Table 5.

Table 4. Design criteria for the surrogate and mixture designs for the acetylene data with $\bar{\xi}^2 = 0.4588$, $\bar{\xi}^2 = 1.0000$.

	$\mathbf{X}$	$\mathbf{X}_k$	$\mathbf{X}_t$	$\mathbf{Z}_t$	$\mathbf{W}_t$
k,t	0	0.0649	0.1240	0.0610	0.2917
VF_M	7682	10.00	10.00	10.00	10.00
κ	47670	52.07	52.07	52.07	53.48
A	15044	39.30	44.87	41.86	52.43
D	17.21	5.78	6.44	6.10	7.17
E	14357	15.38	17.56	16.38	24.15
M	0	0.0108	0.0108	0.0108	0.0102
ρ	0.9968	0.9682	0.9682	0.9682	0.9672
MAD	0	0.0295	0.0289	0.0287	0.0442
Δ	0	0.6171	0.6765	0.6014	1.0148

Table 5. Design criteria for the $\{X_t, Z_t, W_t\}$ designs for the Acetylene Data.

	X_t		Z_t		W_t	
t	A	D	A	D	A	D
0.00	15044	17.21	15044	17.21	15044	17.21
0.20	29.88	5.39	15.02	3.15	95.84	8.48
0.40	16.75	3.98	8.43	1.47	31.88	6.08
0.60	12.31	3.39	6.25	0.61	17.47	4.80
0.80	10.50	3.31	5.30	0.16	12.35	4.12
1.00	10.90	3.90	5.00	0.00	10.90	3.90

Table 6. Design criteria for the surrogate and mixture designs for Body Fat with $\bar{\xi}^2 = 0.6563$ and $\bar{\xi}^2 = 1.0000$.

	X	X_k	X_t	Z_t	W_t
k,t	0	0.0549	0.0773	0.0521	0.2328
VF_M	709	10.00	10.00	10.00	10.00
κ	2844	38.11	21.08	38.11	38.10
A	1378	19.45	19.47	20.52	24.60
D	6.57	2.15	2.39	2.31	2.76
E	1376	17.96	19.47	18.95	22.84
M	0	0.0315	0.0315	0.0315	0.0287
ρ	0.8952	0.8860	0.8860	0.8860	0.8858
MAD	0	0.0258	0.0217	0.0218	0.0241
Δ	0	0.2560	0.2492	0.2240	0.3646

4.3. Body fat data: Neter et al. [20]

The data are given in [20] with $n = 20$and $p = 3$. The explanatory variables are x_1 tricep skinfold thickness; x_2 thigh circumference; x_3 midarm circumference and with y the amount of body fat. The parameters for the three methods have been computed with $VF_M = 10$ to allow for comparisons. All designs show the expected improvement in condition number and design criteria. From Table 6, it is seen that the design $\boldsymbol{Z}_t$ overall indicates the least perturbation of the original design, with values $\Delta = 0.2240$ and nearly the smallest value for $\mathtt{MAD} = 0.0218$. On the other hand, the surrogate design $\boldsymbol{X}_k$ has superior $\{\mathtt{A}, \mathtt{D}, \mathtt{E}\}$ values.

4.4. French economy data

A standard regression analysis is given in [21] to model the French Economy for years 1949–1959 with $n = 11$ and $p = 3$. The variables are x_1 domestic production; x_2 stock formation; x_3 domestic consumption and y imports. From Table 7, it is seen that design $\boldsymbol{Z}_t$ reflects the least perturbation of $\boldsymbol{Z}_t$ from the original design, with smallest $\mathtt{MAD} = 0.0220$ and $\Delta = 0.1921$. On the other hand, the surrogate design $\boldsymbol{X}_k$ has superior $\{\mathtt{A}, \mathtt{D}, \mathtt{E}\}$ values.

4.5. Hospital manpower data

The Hospital Manpower Data comprise records at $n = 17$ U.S. Naval Hospitals with $p = 5$ regressors: x_1 average daily patient load; x_2 monthly X-ray exposures; x_3 monthly occupied bed days; x_4 eligible population in the area divided by 1000; x_5 average length of patients' stay in days and y monthly man-hours as reported in [14]. From Table 8 is seen that the

Table 7. Design criteria for the surrogate and mixture designs for the French Economy with $\bar{\xi}^2 = 0.6751$ and $\bar{\xi}^2 = 1.0000$.

	X	X_k	X_t	Z_t	W_t
k,t	0	0.0513	0.0706	0.0488	0.2050
VF_M	186	10.00	10.00	10.00	10.00
κ	742	38.00	38.00	38.00	38.00
A	373	19.97	21.49	21.00	24.42
D	5.23	2.15	2.37	2.30	2.69
E	371	18.53	19.94	19.48	22.74
M	0	0.0292	0.0292	0.0292	0.0272
ρ	0.9959	0.9948	0.9948	0.9948	0.9948
MAD	0	0.0248	0.0225	0.0220	0.0265
Δ	0	0.2238	0.2170	0.1921	0.3156

Table 8. Design criteria for the surrogate and mixture designs for the Hospital Manpower Data with $\bar{\xi}^2 = 0.4576$ and $\bar{\xi}^2 = 1.0000$.

	X	X_k	X_t	Z_t	W_t
k,t	0	0.0722	0.1363	0.0674	0.2945
VF_M	9598	10.00	10.00	10.00	10.00
κ	77770	59.09	59.09	59.09	64.73
A	18566	30.28	35.05	32.46	40.31
D	14.36	5.45	6.18	5.80	6.59
E	18529	13.84	16.02	14.83	23.93
M	0	0.0286	0.0286	0.0286	0.0214
ρ	0.9954	0.9944	0.9944	0.9944	0.9935
MAD	0	0.0221	0.0231	0.0211	0.0412
Δ	0	0.5571	0.5711	0.5287	0.8910

design $\boldsymbol{Z}_t$ has been perturbed least from the original $\boldsymbol{X}$, having the smallest $\mathtt{MAD} = 0.0211$ and $\Delta = 0.5287$. On the other hand, the surrogate design $\boldsymbol{X}_k$ has superior $\{\mathtt{A}, \mathtt{D}, \mathtt{E}\}$ values.

4.6. Number of active metropolitan physicians

We use the Standard Metropolitan Statistical Area (SMSA) data having $n = 141$ and $p = 3$ from the website.[1] The variables are x_1 total population (in thousands); x_2 land area (in square miles); x_3 total personal income (in millions of dollars) and y number of active physicians. From Table 9, the design $\boldsymbol{Z}_t$ reflects the least perturbation from the original design, having the smallest values $\mathtt{MAD} = 0.1969$ and $\Delta = 0.0047$. On the other hand, the surrogate design $\boldsymbol{X}_k$ has superior $\{\mathtt{A}, \mathtt{D}, \mathtt{E}\}$ values.

5. Conclusion

5.1. Summary

This study advances a new class of linear estimators as mixtures in efforts to mitigate collinearity. The procedure is based on mixing the original design with a minimal design, or mixing its Fisher information matrix with a scalar matrix as target, giving the ensembles $\{\boldsymbol{X}_t, \boldsymbol{Z}_t, \boldsymbol{W}_t;\ t \in [0,1]\}$. Theorem 3.1 shows that $\mathtt{MSE}(\hat{\boldsymbol{\beta}}_\omega(t))$ is decreasing at $t = 0$ for each of $\{\boldsymbol{X}_t, \boldsymbol{Z}_t, \boldsymbol{W}_t;\ t \in [0,1]\}$, so that the solutions are admissible and thus well-conditioned alternatives to $\mathtt{OLS}$. Theorem 3.2 establishes that the $(\kappa, \mathtt{A}, \mathtt{D}, \mathtt{E})$-efficiency indices for

Table 9. Design criteria for the surrogate and mixture designs for the SMSA data with $\bar{\xi}^2 = 0.6707$ and $\bar{\xi}^2 = 1.0000$.

	X	X_k	X_t	Z_t	W_t
k,t	0	0.0518	0.0717	0.0492	0.2087
VF_M	211	10.00	10.00	10.00	10.00
κ	854	38.48	38.48	38.48	38.44
A	423	19.93	21.47	20.96	24.48
D	5.37	2.16	2.39	2.32	2.71
E	420	18.47	19.89	19.42	22.78
M	0	0.0297	0.0297	0.0297	0.0276
ρ	0.9789	0.9787	0.9787	0.9787	0.9787
MAD	0	0.0052	0.0049	0.0047	0.0058
Δ	0	0.2280	0.2209	0.1969	0.3214

Table 10. Summary comparing surrogate (X_k) and mixture (Z_t) designs.

Case study	MAD(X_k)	MAD(Z_t)	%	$\Delta(X_k)$	$\Delta(Z_k)$	%
Acetylene	0.0295	0.0287	2.7	0.6171	0.6014	2.5
Body fat	0.0258	0.0218	15.5	0.2560	0.2240	12.5
French economy	0.0248	0.0220	11.3	0.2238	0.1921	14.2
Hospital	0.0221	0.0211	4.5	0.5571	0.5287	5.1
SMSA	0.0052	0.0047	9.6	0.2280	0.1969	13.6

$\hat{\boldsymbol{\beta}}(W_t)$ decrease monotonically for $t \uparrow \in [0,1]$ viewed as functions on $(\mathbb{F}^{\tau}_{n\times p}, \succeq)$; and similarly, the $(\kappa, \mathrm{A}, \mathrm{D}, \mathrm{E})$ indices for $\hat{\boldsymbol{\beta}}(Z_t)$ decrease monotonically for $t \uparrow \in [0,1]$ viewed as functions on $(\mathbb{S}^+_p, \succeq_L)$. For a fixed $t \in [0,1]$, it is shown in Theorem 3.3 that the Fisher information matrices are Loewner ordered as $\{W_t'W_t \preceq_L X_t'X_t \preceq_L Z_t'Z_t\}$, and their inverses as $\{\boldsymbol{\Sigma}_W \succeq_L \boldsymbol{\Sigma}_X \succeq_L \boldsymbol{\Sigma}_Z\}$, the latter as dispersion matrices for the corresponding $\hat{\boldsymbol{\beta}}_\omega$, and thus each efficiency index $(\kappa, \mathrm{A}, \mathrm{D}, \mathrm{E})$ is ordered as are $\{\boldsymbol{\Sigma}_W \succeq_L \boldsymbol{\Sigma}_X \succeq_L \boldsymbol{\Sigma}_Z\}$ in $(\mathbb{S}^+_p, \succeq_L)$.

5.2. Comparing the mixtures

In comparisons among designs, we seek a balance between (i) efficiency and (ii) proximity to the original design, since highly ill-conditioned data often stem from constrained ranges of the settings. Thus modified designs might reflect those same constraints. In addition, design points so discovered may suggest improved yet feasible design points in subsequent experiments. In this regard, the MAD and Δ diagnostics may be especially helpful to users.In retrospect, each case study demonstrated Z_t to be nearer to the original design in both MAD and Δ. In Table 10 are summarized advantages of Z_t over the surrogate and ridge designs. The relative improvement is shown in the percentage column identified with %, in some cases negligible. Again k and t achieve $VF_M = 10$.

Details regarding efficiencies are summarized in Table 11. As noted, the surrogate design X_k dominates in the (A, D, E) criteria, often negligibly, in comparison with Z_t. On the other hand, W_t exhibits somewhat larger values. Such diagnostics may serve to inform users regarding the tradeoff between efficiency and proximity to the original design, in all cases improving uniformly over designs known to be excessively ill-conditioned.

Surrogate and ridge solutions offer OLS-admissibility for $k \in [0, \infty)$ and $k \in [0, k_0)$, respectively, yet remain somewhat equivocal as to those diverse choices. On the other hand,

Table 11. Comparison of (A, D, E) efficiency indices across choices among $(\boldsymbol{X}_k, \boldsymbol{X}_t, \boldsymbol{Z}_t, \boldsymbol{W}_t)$ for the five cases studies.

$\boldsymbol{X}_k$			$\boldsymbol{X}_t$			$\boldsymbol{Z}_t$			Wt		
A	D	E	A	D	E	A	D	E	A	D	E
Acetylene											
39.30	5.78	15.38	44.87	6.44	17.56	41.86	6.10	16.38	52.43	7.17	24.15
Body fat											
19.45	2.15	17.93	19.47	2.39	19.47	20.52	2.31	18.95	24.60	2.76	22.84
French economy											
19.97	2.15	18.53	21.49	2.37	19.94	21.00	2.30	19.48	24.42	2.69	22.74
Hospital											
30.28	5.45	13.84	35.05	6.18	16.02	32.46	5.80	14.83	40.31	6.59	23.93
SMSA											
19.93	2.16	18.47	21.47	2.39	19.89	20.96	2.32	19.42	24.48	2.71	22.78

Table 12. Design criteria for the surrogate–mixture designs for body fat with $k = 0.0549$.

t	VF_M	κ	A	D	E
0.0	10.0000	38.1080	19.45	2.15	17.96
0.2	2.3952	7.4675	5.44	0.72	3.91
0.4	1.4958	3.7218	3.77	0.24	2.20
0.6	1.1752	2.2611	3.17	0.00	1.53
0.8	1.0392	1.4832	2.92	−0.12	1.17
1.0	1.0000	1.0000	2.84	−0.16	0.95

the mixture models serve to pull a design or its Fisher information matrix towards perfectly conditioned, albeit unattainable, targets. It is apparent that mixtures apply also in surrogate models, beginning now with $\boldsymbol{X}_k$ instead of $\boldsymbol{X}$. Accordingly, to simplify notation, denote by $\sigma(\boldsymbol{X}_k) = \{\xi_1, \ldots, \xi_p\}$, clearly depending on k. As an example, consider the Body Fat Data of Section 4.3 with surrogate design $\boldsymbol{X}_k$ having $k = 0.0549$. Working towards Table 12, take $\boldsymbol{Z}_t$ from Table 3 as $\boldsymbol{T}_t$ to get the surrogate mixtures $\boldsymbol{T}_t = \boldsymbol{P}\,\mathrm{Diag}([(1-t)\xi_i^2 + t\bar{\xi^2}]^{1/2}\boldsymbol{Q}'$ and $\boldsymbol{T}_t{}'\boldsymbol{T}_t = \boldsymbol{Q}\,\mathrm{Diag}([(1-t)\xi_i^2 + t\bar{\xi^2}])\boldsymbol{Q}'$ with $\bar{\xi^2} = 1.05494$. The ensemble $\{\boldsymbol{T}_t;\ t \in [0,1]\}$ thus varies from the surrogate design $\boldsymbol{X}_k$ to an orthogonal design. Table 12 reports the efficiency indices for the family $\{\boldsymbol{T}_t;\ t \in [0,1]\}$ for varying values of the mixing parameter t, demonstrating how the surrogate design can be further enhanced through mixtures.

5.3. Performance of the algorithms

The several case studies enable a preliminary assessment as to the performance of our algorithms. Details are summarized in Table 13, where the case studies are arranged in order of decreasing $\kappa(\boldsymbol{X})$ in the original data. It is seen that choices for k in surrogate regression, namely $k(\boldsymbol{X}_k)$, decrease monotonically with decreasing values of $\kappa(\boldsymbol{X})$. In like manner, the choice for $t \in [0,1]$ is seen to be monotone decreasing with decreasing $\kappa(\boldsymbol{X})$ for each of $\{\boldsymbol{X}_t, \boldsymbol{Z}_t, \boldsymbol{W}_t\}$. In addition, from their definitions in Tables 2 and 3, where $\boldsymbol{W}_t$ adjusts first-order effects and $(\{\boldsymbol{X}_k, \boldsymbol{X}_t, \boldsymbol{Z}_t)$ adjust moments of second order, it is plausible that $[t(\boldsymbol{W}_t)]^2$ should approximate both $k(\boldsymbol{X}_k)$ and $t(\boldsymbol{Z}_t)$, which is supported in Table 13. In summary, our algorithms are seen to perform consistently over a wide range of ill-conditioned data.

Table 13. Summary comparing performance of the algorithms in use.

Case study	$\kappa(\boldsymbol{X})$	$k(\boldsymbol{X}_k)$	$t(\boldsymbol{X}_t)$	$t(\boldsymbol{Z}_t)$	$t(\boldsymbol{W}_t)$	$[t(\boldsymbol{W}_t)]^2$
Hospital	77770	0.0722	0.1363	0.0674	0.2945	0.0867
Acetylene	47670	0.0649	0.1240	0.0610	0.2917	0.0851
Body fat	2844	0.0549	0.0773	0.0521	0.2328	0.0542
SMSA	854	0.0518	0.0717	0.0492	0.2087	0.0436
French economy	742	0.0513	0.0706	0.0488	0.2050	0.0420

Table 14. Design criteria for the surrogate and mixture models for the SMSA data with $VF_M = 5$.

	$\boldsymbol{X}$	$\boldsymbol{X}_k$	$\boldsymbol{X}_t$	$\boldsymbol{Z}_t$	$\boldsymbol{W}_t$
k,t	0	0.1158	0.1473	0.1038	0.3133
VF_M	211	5.00	5.00	5.00	5.00
κ	854	18.17	18.17	18.17	18.14
A	423	9.85	11.55	10.99	13.69
D	5.37	1.29	1.77	1.62	2.19
E	420	8.46	9.92	9.44	11.88
M	0	0.1029	0.1029	0.1029	0.0944
ρ	0.9789	0.9787	0.9787	0.9787	0.9786
MAD	0	0.0085	0.0077	0.0074	0.0088
Δ	0	0.3924	0.3632	0.3166	0.4826

5.4. Choice of tuning parameters

Recalling $VF_M = \max\{VIF(\hat{\beta}_j); 1 \le i \le p\}$, we have followed the common rule-of-thumb that ill-conditioning occurs when $VF_M \ge 10$, and accordingly have chosen the tuning parameters, either k or t, to satisfy $VF_M = 10$. As this is arbitrary, we return to the SMSA data in Section 4.6 and set $VF_M = 5$ as the benchmark. The results are given in Table 14; these show once again that the altered $\boldsymbol{Z}_t$ reflects the least perturbation from the original design with MAD $= 0.0074$ and $\Delta = 0.3166$; and that the surrogate $\mathbf{X}_k$ has superior (A, D, E) efficiencies.

5.5. In retrospect

Our original goals in seeking alternatives to ridge regression were to overcome two problems, namely, (1) the condition number κ for the dispersion matrix $\mathrm{V}(\hat{\boldsymbol{\beta}}_R(k))$ is not monotone in k but tends back to that of $\mathrm{V}(\hat{\boldsymbol{\beta}}_R(0))$ for the original OLS solution and (2) for large k the ridge model becomes infeasible, having an infinite moment matrix and solutions $\hat{\boldsymbol{\beta}}_R(k) \to \mathbf{0}$. For surrogate regression, Jensen and Ramirez [12] established that the condition number κ for $\mathrm{V}(\hat{\boldsymbol{\beta}}_S(k))$ is indeed monotone and tends to 1. The condition number facts follow on noting at $\sigma^2 = 1$ that the eigenvalues of $\mathrm{V}(\hat{\boldsymbol{\beta}}_S(k))$ are $\{1/(\xi_i^2 + k)\}$ and of $\mathrm{V}(\hat{\boldsymbol{\beta}}_R(k))$ are $\{\xi_i^2/(\xi_i^2 + k)^2\}$ with $\{1 \le i \le p\}$, so that

$$\lim_{k\to\infty} \kappa[\mathrm{V}(\hat{\boldsymbol{\beta}}_R(k)] = \lim_{k\to\infty} \frac{\xi_1^2}{\xi_p^2} \frac{(\xi_p^2 + k)^2}{(\xi_1^2 + k)^2} = \frac{\xi_1^2}{\xi_p^2} = \kappa[\mathrm{V}(\hat{\boldsymbol{\beta}}_L)]$$

for OLS as claimed. In addition, both $\mathrm{V}(\hat{\boldsymbol{\beta}}_S(k))$ and $\mathrm{V}(\hat{\boldsymbol{\beta}}_R(k))$ tend to zero in the Frobenius matrix norm. These considerations solve problem (1).

To avoid problem (2), the mixtures $\{\boldsymbol{X}_t;\ t \in [0, 1]\}$ were introduced in [4] as noted. The present study introduces two additional mixtures, namely, $\boldsymbol{Z}_t$ and $\boldsymbol{W}_t$, the first by mixing

the eigenvalues of the original moment matrix with their average, and the second by mixing the singular values of the original design with their average. These procedures both serve to circumvent problems (1) and (2). For our case studies, $\boldsymbol{Z}_t$ was seen to be the superior procedure. A forthcoming study will undertake further extensions such as mixing the eigenvalues using the geometric means *in lieu of* the arithmetic means of the present study.

Note

1. https://onlinecourses.science.psu.edu/stat857/sites/onlinecourses.science.psu.edu.stat857/files/smsa.data

Disclosure statement

No potential conflict of interest was reported by the authors.

ORCID

D. E. Ramirez http://orcid.org/0000-0002-2419-4538

References

[1] Hadi AS. Ridge and surrogate ridge regressions. In: Lovric M, editor. International encyclopedia of statistical science. Berlin: Springer; 2011.
[2] Hoerl AE, Kennard RW. Ridge regression: biased estimation for nonorthogonal problems. Technometrics. 1970;12:55–67.
[3] Jensen DR, Ramirez DE. Anomalies in the foundations of ridge regression. Int Stat Rev. 2008;76:89–105.
[4] Jensen DR, Ramirez DE. Singular majorants and minorants: enhanced design conditioning. J Stat Comput Simul. 2017;87:1827–1841.
[5] Marshall AW, Olkin I. Inequalities: theory of majorization and its applications. New York: Academic Press; 1979.
[6] Loewner C. Über monotone Matrixfunktonen. Math Z. 1934;38:177–216.
[7] Jensen DR. Matrix extremes and related stochastic bounds. In: Shaked M, Tong TL, editors. Stochastic inequalities. Hayward (CA): Institute of Mathematical Statistics; 1992. p. 133–144.
[8] Jensen DR. Invariant ordering and order preseration. In: Tong YL, editor. Inequalities in statistics and probability. Hayward (CA): Institute of Mathematical Statistics; 1984. p. 26–34.
[9] Jensen DR. Condition numbers and D-efficiency. Stat Probab Lett. 2004;66:267–274.
[10] Jensen DR, Ramirez DE. Enhanced design efficiency through least upper bounds. J Stat Comput Simul. 2016;86:1798–1817.
[11] Marquardt DW, Snee RD. Ridge regression in practice. Amer Statist. 1975;29:3–20.
[12] Jensen DR, Ramirez DE. Surrogate models in ill-conditioned systems. J Statist Plan Inference. 2010;140:2069–2077.
[13] Jensen DR, Ramirez DE. Revision: variance inflation in regression. Adv Decis Sci. 2012;2012:1–15.
[14] Myers RH. Classical and modern regression with applications. 2nd ed. Boston, MA: PWS-Kent Publishing; 1990.
[15] O'Brien RM. A caution regarding rules of thumb for variance inflation factors. Qual Quant. 2007;41:673–690.
[16] Sengupta D, Bhimasankaram P. On the roles of observations in collinearity in the linear model. J Amer Statist Assoc. 1997;92:1024–1032.
[17] Jensen DR, Ramirez DE. Tracking MSE efficiencies in ridge regression. Adv Appl Stat Sci. 2010;1:381–398.

[18] Cvetkovski Z. Inequalities: theorems, techniques and selected problems. New York: Springer; 2012.
[19] Mauchly JW. Significance test for sphericity of a normal n-variate distribution. Ann Math Statist. 1940;11:204–209.
[20] Neter J, Kutner MH, Nachtsheim CJ, et al. Applied linear statistical models. 4th ed. Chicago: Irwin; 1996.
[21] Chatterjee S, Hadi A. Regression analysis by example. 4th ed. Hoboken, NJ: John Willey & Sons; 2006.