Supplementary material for ‘Minimax efficient random experimental design strategies with application to model-robust design for prediction’

Timothy W. Waite[*] and David C. Woods[†]

December 9, 2020

A Background: games and their solution

In a two-person zero-sum game, Players I and II take actions $\boldsymbol{\theta} \in \Theta$ and $\boldsymbol{\xi} \in \Xi$, respectively. Given these choices, Player II experiences a loss $\mathcal{L}(\boldsymbol{\theta}, \boldsymbol{\xi})$, which they wish to minimize, and Player I experiences an equal gain or ‘payoff’, which they wish to maximize; both players must account for the possible decisions of the other player. A player’s actions may be chosen according to a random strategy defined by a probability measure on their action space, Θ or Ξ, as appropriate. Statistical decision problems and experimental design may be viewed as a game in which Nature corresponds to Player I and the Statistician corresponds to Player II (Berger 1985, p.347, Wu 1981). In experimental design, $\mathcal{L}(\boldsymbol{\theta}, \boldsymbol{\xi}) = R(\boldsymbol{\theta}, \boldsymbol{\xi}) = \mathrm{E}_{\mathbf{y}|\boldsymbol{\theta},\boldsymbol{\xi}}\, \ell(\boldsymbol{\theta}, \hat{\boldsymbol{\alpha}})$. Let π be a random strategy for Player II. When the action $\boldsymbol{\xi}$ is chosen, $\boldsymbol{\theta}$ is not known: to address this the minimax strategy minimizes $\max_{\boldsymbol{\theta}' \in \Theta} \mathcal{L}(\boldsymbol{\theta}', \pi)$, where $\mathcal{L}(\boldsymbol{\theta}', \pi) = \mathrm{E}_{\mathbf{y},\boldsymbol{\xi}|\boldsymbol{\theta}'}\, \ell(\boldsymbol{\theta}', \hat{\boldsymbol{\alpha}}) = R(\boldsymbol{\theta}', \pi)$.

For any game in which Players I and II each have available finitely many actions, $\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_M$ and $\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_K$, respectively, the loss function can be recorded in an $M \times K$

[*]Department of Mathematics, University of Manchester, UK
[†]Statistical Sciences Research Institute, University of Southampton, UK

matrix $\mathbf{L}$, with (m,k)th entry $\mathcal{L}(\boldsymbol{\theta}_m, \boldsymbol{\xi}_k)$, $m = 1, \ldots, M$, $k = 1, \ldots, K$. Moreover, a random strategy for Player II can be described by a vector $\boldsymbol{\pi} = [\pi(\boldsymbol{\xi}_1), \ldots, \pi(\boldsymbol{\xi}_K)]^{\mathrm{T}}$ specifying the probability, $\pi(\boldsymbol{\xi}_k)$, of selecting action $\boldsymbol{\xi}_k$, $k = 1, \ldots, K$. Provided all entries of the matrix $\mathbf{L}$ are positive, a minimax optimal strategy $\boldsymbol{\pi}_{\mathrm{mM}} = [\pi_{\mathrm{mM}}(\boldsymbol{\xi}_1), \ldots, \pi_{\mathrm{mM}}(\boldsymbol{\xi}_K)]^{\mathrm{T}}$ for Player II can be computed as $\boldsymbol{\pi}_{\mathrm{mM}} = \boldsymbol{\pi}^\dagger / [\mathbf{1}_K^{\mathrm{T}} \boldsymbol{\pi}^\dagger]$, where $\boldsymbol{\pi}^\dagger$ is a solution to the linear program

$$\text{Maximize } \mathbf{1}_K^{\mathrm{T}} \boldsymbol{\pi}^\dagger \text{ subject to } \mathbf{L}\boldsymbol{\pi}^\dagger \leq \mathbf{1}_M \,, \boldsymbol{\pi}^\dagger \geq \mathbf{0} \,, \tag{A.1}$$

with $\mathbf{1}_K = (1, 1, \ldots, 1)^{\mathrm{T}}$ a K-vector, and vector inequalities defined entrywise (e.g. Thie & Keough 2011, p.363). Above, $\boldsymbol{\pi}^\dagger$ may be obtained using Dantzig's simplex algorithm, implemented for example in `lpsolve` (Berkelaar et al. 2004), accessible via the R package `lpSolveAPI` (Konis 2014). A minimax strategy may also be obtained with the same method if $\mathbf{L}$ has non-positive entries via the addition of a constant to all elements of the matrix to ensure positivity, since this operation does not affect the solution space. Similar considerations apply for Player I. We apply the above to find optimal random designs for discrete design spaces in Section 4.

B Proofs of analytical results

B.1 Further details of results in Section 3

B.1.1 Proofs of analytical results

Proof of Proposition 3.1 Recall that $\sum_{k=1}^n \epsilon_k(\mathbf{x}) = 0$ for all $\mathbf{x} \in \mathcal{X}$. Hence the unit effect term for the $\rho(i)$th unit, $\rho \sim \mathrm{Uniform}(S_n)$, satisfies

$$\mathrm{E}[\epsilon_{\rho(i)}(\tilde{\mathbf{x}}_i)] = \frac{1}{n!} \sum_{\rho \in S_n} \epsilon_{\rho(i)}(\tilde{\mathbf{x}}_i) = \frac{1}{n!} \sum_{i'=1}^{n} \sum_{\rho \in S_n : \rho(i) = i'} \epsilon_{\rho(i)}(\tilde{\mathbf{x}}_i) = \frac{1}{n} \sum_{i'=1}^{n} \epsilon_{i'}(\tilde{\mathbf{x}}_i) = 0 \,,$$

and $\mathrm{Var}[\epsilon_{\rho(i)}(\tilde{\mathbf{x}}_i)] = \mathrm{E}[\epsilon_{\rho(i)}(\tilde{\mathbf{x}}_i)^2] = \frac{1}{n} \sum_{i'=1}^n \epsilon_{i'}(\tilde{\mathbf{x}}_i)^2 = S^2(\tilde{\mathbf{x}}_i)$. Furthermore, for $i \neq j$, we have

$$\mathrm{Cov}[\epsilon_{\rho(i)}(\tilde{\mathbf{x}}_i), \epsilon_{\rho(j)}(\tilde{\mathbf{x}}_j)] = \mathrm{E}[\epsilon_{\rho(i)}(\tilde{\mathbf{x}}_i) \epsilon_{\rho(j)}(\tilde{\mathbf{x}}_j)] = \frac{1}{n!} \sum_{\rho \in S_n} \epsilon_{\rho(i)}(\tilde{\mathbf{x}}_i) \epsilon_{\rho(j)}(\tilde{\mathbf{x}}_j)$$

$$= \frac{1}{n!} \sum_{(k,l):k\neq l} \sum_{\rho\in S_n:\rho(i)=k,\rho(j)=l} \epsilon_{\rho(i)}(\tilde{\mathbf{x}}_i)\epsilon_{\rho(j)}(\tilde{\mathbf{x}}_j) = \frac{1}{n(n-1)} \sum_{(k,l):k\neq l} \epsilon_k(\tilde{\mathbf{x}}_i)\epsilon_l(\tilde{\mathbf{x}}_j)$$

$$= \frac{1}{n(n-1)} \left[\sum_{(k,l)} \epsilon_k(\tilde{\mathbf{x}}_i)\epsilon_l(\tilde{\mathbf{x}}_j) - \sum_k \epsilon_k(\tilde{\mathbf{x}}_i)\epsilon_k(\tilde{\mathbf{x}}_j) \right]$$

$$= \frac{1}{n(n-1)} \left[\sum_k \epsilon_k(\tilde{\mathbf{x}}_i) \sum_l \epsilon_l(\tilde{\mathbf{x}}_j) - \sum_k \epsilon_k(\tilde{\mathbf{x}}_i)\epsilon_k(\tilde{\mathbf{x}}_j) \right] = -\frac{1}{n(n-1)} \boldsymbol{\epsilon}(\tilde{\mathbf{x}}_i)^{\mathrm{T}} \boldsymbol{\epsilon}(\tilde{\mathbf{x}}_j) \,.$$

Since $r_i = Y_{\rho(i)}(\tilde{\mathbf{x}}_i) = \mathbf{f}^{\mathrm{T}}(\tilde{\mathbf{x}}_i)\boldsymbol{\beta} + \epsilon_{\rho(i)}(\tilde{\mathbf{x}}_i)$, we have that $\mathrm{E}[r_i] = \mathbf{f}^{\mathrm{T}}(\tilde{\mathbf{x}}_i)\boldsymbol{\beta} + \mathrm{E}[\epsilon_{\rho(i)}(\tilde{\mathbf{x}}_i)] = \mathbf{f}^{\mathrm{T}}(\tilde{\mathbf{x}}_i)\boldsymbol{\beta}$, so $\mathrm{E}[\mathbf{r}] = \mathbf{F}_{\tilde{\boldsymbol{\xi}}}\boldsymbol{\beta}$, giving the result for the expectation. For the variance, note that $\mathrm{Var}[r_i] = \mathrm{Var}[\epsilon_{\rho(i)}(\tilde{\mathbf{x}}_i)] = S^2(\tilde{\mathbf{x}}_i)$ and $\mathrm{Cov}(r_i, r_j) = -\frac{1}{n(n-1)}\boldsymbol{\epsilon}(\tilde{\mathbf{x}}_i)^{\mathrm{T}}\boldsymbol{\epsilon}(\tilde{\mathbf{x}}_j)$.

The last step is to verify that the (i,j)th entry of

$$\mathbf{V}(\tilde{\boldsymbol{\xi}}) = \frac{n}{n-1} \operatorname{diag}[S^2(\tilde{\mathbf{x}}_1), \ldots, S^2(\tilde{\mathbf{x}}_n)] - \frac{1}{n(n-1)} \mathbf{E}^{\mathrm{T}}(\tilde{\boldsymbol{\xi}})\mathbf{E}(\tilde{\boldsymbol{\xi}})$$

coincides with the (i,j)th entry of $\mathrm{Var}(\mathbf{r})$ for all (i,j). For $i \neq j$, we have that $[\mathbf{V}(\tilde{\boldsymbol{\xi}})]_{ij} = -\frac{1}{n(n-1)}\boldsymbol{\epsilon}(\tilde{\mathbf{x}}_i)^{\mathrm{T}}\boldsymbol{\epsilon}(\tilde{\mathbf{x}}_j) = \mathrm{Cov}(r_i, r_j)$. We also have $[\mathbf{V}(\tilde{\boldsymbol{\xi}})]_{ii} = \frac{n}{n-1}S^2(\tilde{\mathbf{x}}_i) - \frac{1}{n(n-1)}\boldsymbol{\epsilon}^{\mathrm{T}}(\tilde{\mathbf{x}}_i)\boldsymbol{\epsilon}(\tilde{\mathbf{x}}_i) = \frac{n}{n-1}S^2(\tilde{\mathbf{x}}_i) - \frac{1}{n-1}S^2(\tilde{\mathbf{x}}_i) = S^2(\tilde{\mathbf{x}}_i) = \mathrm{Var}(r_i)$. The result is proved.

Lemma B.1. *Suppose that $\mathcal{X}$ is compact and the $f_j : \mathcal{X} \to \mathbb{R}$, $(j = 0, 1 \ldots, p)$, are continuous with respect to the topology on $\mathcal{X}$. If the RDS π is non-singular then (i) $\det \mathbf{M}_{\mathbf{d}}$ is bounded away from zero for $\mathbf{d} \in \mathrm{supp}(\pi)$, and (ii) the map $\mathbf{d} \mapsto \mathbf{M}_{\mathbf{d}}^{-1}$ is continuous on $\mathrm{supp}(\pi)$.*

Proof. Statement (i) is proved by contradiction. Suppose instead that $\det \mathbf{M}_{\mathbf{d}}$ is not bounded away from zero. Then there exists a sequence $\mathbf{d}_1, \mathbf{d}_2, \ldots \in \mathrm{supp}(\pi)$ with $\det \mathbf{M}_{\mathbf{d}_s} \to 0$ as $s \to \infty$. By standard properties of the support, $\mathrm{supp}(\pi)$ is a closed subset of the compact Euclidean space $\mathcal{X}^n$, hence it is sequentially compact. Thus there exists a subsequence $\mathbf{d}_{s_1}, \mathbf{d}_{s_2}, \ldots$ converging to a limit $\mathbf{d}^* \in \mathrm{supp}(\pi)$.

Note that the map $\mathbf{d} \mapsto \mathbf{M}_{\mathbf{d}}$ is continuous on $\mathrm{supp}(\pi)$, as $[\mathbf{M}_{\mathbf{d}}]_{jk} = \sum_{i=1}^n f_j(\mathbf{x}_i) f_k(\mathbf{x}_i)$ and the f_j are continuous. Hence the function $g : \mathrm{supp}(\pi) \to \mathbb{R}$ defined by $g(\mathbf{d}) = \det \mathbf{M}_{\mathbf{d}}$

is also continuous. Therefore we must have that $\det \mathbf{M}_{\mathbf{d}^*} = g(\mathbf{d}^*) = \lim_{h\to\infty} g(\mathbf{d}_{s_h}) = \lim_{h\to\infty} \det \mathbf{M}_{\mathbf{d}_{s_h}} = 0$. Hence $\mathbf{d}^*$ is an element of $\mathrm{supp}(\pi)$ with $\det \mathbf{M}_{\mathbf{d}^*} = 0$, contradicting the assumption that the RDS is singular.

Part (ii) then follows from the fact that $\mathbf{M}_{\mathbf{d}}^{-1} = \frac{1}{\det \mathbf{M}_{\mathbf{d}}} \operatorname{adj} \mathbf{M}_{\mathbf{d}}$ is a product of continuous functions.

□

Before proving Theorem 3.2 we need some preliminary notation and results. Let $\boldsymbol{\epsilon}(\boldsymbol{\xi}) = (\epsilon_1(\mathbf{x}_1), \ldots, \epsilon_n(\mathbf{x}_n))^{\mathrm{T}}$, noting that given $\boldsymbol{\xi}$ and $\boldsymbol{\theta} = (\boldsymbol{\beta}, \boldsymbol{\epsilon}) \in \Theta = \mathbb{R}^{p+1} \times \mathcal{E}$ the response vector is determined as $\mathbf{y} = \boldsymbol{\gamma}_{\boldsymbol{\theta},\boldsymbol{\xi}} = \mathbf{F}_{\boldsymbol{\xi}}\boldsymbol{\beta} + \boldsymbol{\epsilon}(\boldsymbol{\xi})$. Given $\rho \in S_n$, $\mathbf{v} \in \mathbb{R}^n$, $\mathbf{B} = [b_{ij}] \in \mathbb{R}^{n\times k}$, $\boldsymbol{\epsilon} \in \mathcal{E}$, $\boldsymbol{\beta} \in \mathbb{R}^{p+1}$ and $\boldsymbol{\theta} = (\boldsymbol{\beta}, \boldsymbol{\epsilon})$, define $\rho(\mathbf{v})$ by $[\rho(\mathbf{v})]_i = \mathbf{v}_{\rho^{-1}(i)}$,

$$\rho(\mathbf{B}) = \begin{pmatrix} b_{\rho^{-1}(1)1} & \ldots & b_{\rho^{-1}(1)k} \\ \vdots & \ddots & \vdots \\ b_{\rho^{-1}(n)1} & \ldots & b_{\rho^{-1}(n)k} \end{pmatrix},$$

$\rho(\boldsymbol{\theta}) = (\boldsymbol{\beta}, \rho(\boldsymbol{\epsilon}))$, and $\rho(\boldsymbol{\epsilon}) = (\epsilon_{\sigma^{-1}(1)}, \ldots, \epsilon_{\sigma^{-1}(n)})$. The latter is the vector function obtained by permuting the component functions of $\boldsymbol{\epsilon}$. Note that by our assumptions $\rho(\boldsymbol{\epsilon}) \in \mathcal{E}$ and $\rho(\mathcal{E}) = \mathcal{E}$, i.e. $\mathcal{E}$ is invariant under ρ. Also $\rho(\Theta) = \Theta$.

Lemma B.2. *For a design realization $\boldsymbol{\xi} \in \Xi$, $\boldsymbol{\epsilon} \in \mathcal{E}$, and $\rho \in S_n$, we have*

$$\rho(\rho^{-1}(\boldsymbol{\epsilon})(\boldsymbol{\xi})) = \boldsymbol{\epsilon}(\rho(\boldsymbol{\xi})) \tag{B.1}$$

and $\boldsymbol{\gamma}_{\boldsymbol{\theta},\rho(\boldsymbol{\xi})} = \rho(\boldsymbol{\gamma}_{\rho^{-1}(\boldsymbol{\theta}),\boldsymbol{\xi}})$.

Proof of Lemma B.2 Note that $[\rho^{-1}(\boldsymbol{\epsilon})(\boldsymbol{\xi})]_{i'} = \epsilon_{\rho(i')}(\mathbf{x}_{i'})$ and so setting $i' = \rho^{-1}(i)$ we have that the ith component of the left hand side of (B.1) is

$$\begin{aligned}[\rho(\rho^{-1}(\boldsymbol{\epsilon})(\boldsymbol{\xi}))]_i &= [\rho^{-1}(\boldsymbol{\epsilon})(\boldsymbol{\xi})]_{\rho^{-1}(i)} = \epsilon_{\rho(\rho^{-1}(i))}(\mathbf{x}_{\rho^{-1}(i)}) \\ &= \epsilon_i(\mathbf{x}_{\rho^{-1}(i)}) = \epsilon_i([\rho(\boldsymbol{\xi})]_i) = [\boldsymbol{\epsilon}(\rho(\boldsymbol{\xi}))]_i\,.\end{aligned}$$

For the second part, note using (B.1) that $\boldsymbol{\gamma}_{\boldsymbol{\theta},\rho(\boldsymbol{\xi})} = \mathbf{F}_{\rho(\boldsymbol{\xi})}\boldsymbol{\beta}+\boldsymbol{\epsilon}(\rho(\boldsymbol{\xi})) = \rho(\mathbf{F}_{\boldsymbol{\xi}})\boldsymbol{\beta}+\rho(\rho^{-1}(\boldsymbol{\epsilon})(\boldsymbol{\xi})) = \rho(\mathbf{F}_{\boldsymbol{\xi}}\boldsymbol{\beta} + \rho^{-1}(\boldsymbol{\epsilon})(\boldsymbol{\xi})) = \rho(\boldsymbol{\gamma}_{\rho^{-1}(\boldsymbol{\theta}),\boldsymbol{\xi}})$. This is enough to establish the lemma.

Lemma B.3. *Suppose that for each* $\boldsymbol{\xi}' \in \operatorname{supp}(\pi)$, $\mathbf{C}_{\boldsymbol{\xi}'}$ *is a* $k \times (p+1)$ *matrix, and that the map* $\boldsymbol{\xi}' \mapsto \mathbf{C}_{\boldsymbol{\xi}'}$ *is continuous on* $\operatorname{supp}(\pi)$. *If* $\boldsymbol{\beta}^{\mathrm{T}}\,\mathrm{E}[\mathbf{C}_{\boldsymbol{\xi}}^{\mathrm{T}}\mathbf{C}_{\boldsymbol{\xi}}]\boldsymbol{\beta} = 0$ *for all* $\boldsymbol{\beta} \in \mathbb{R}^{p+1}$, *then* $\mathbf{C}_{\boldsymbol{\xi}'} = \mathbf{0}$ *for all* $\boldsymbol{\xi}' \in \operatorname{supp}(\pi)$.

Proof of Lemma B.3 Suppose the contrary, i.e. that there exists $\boldsymbol{\xi}' \in \operatorname{supp}(\pi)$ such that $\mathbf{C}_{\boldsymbol{\xi}'} \neq \mathbf{0}$. Then there also exists some $\boldsymbol{\beta}' \in \mathbb{R}^{p+1}$ such that $\boldsymbol{\beta}'^{\mathrm{T}}\mathbf{C}_{\boldsymbol{\xi}'}^{\mathrm{T}}\mathbf{C}_{\boldsymbol{\xi}'}\boldsymbol{\beta}' > 0$. Moreover, by the assumed continuity of $\mathbf{C}_{\boldsymbol{\xi}}$ there exists an $\epsilon > 0$ together with an open neighbourhood U, $\boldsymbol{\xi}' \in U \subseteq \mathcal{X}^n$, such that $\boldsymbol{\beta}'^{\mathrm{T}}\mathbf{C}_{\boldsymbol{\xi}''}^{\mathrm{T}}\mathbf{C}_{\boldsymbol{\xi}''}\boldsymbol{\beta}' > \epsilon$ for $\boldsymbol{\xi}'' \in U$. By the definition of a support, since $\boldsymbol{\xi}' \in \operatorname{supp}(\pi)$ every open neighbourhood of $\boldsymbol{\xi}'$ has positive measure. Thus $\pi(U) > 0$ and

$$\mathrm{E}[\boldsymbol{\beta}'^{\mathrm{T}}\mathbf{C}_{\boldsymbol{\xi}}^{\mathrm{T}}\mathbf{C}_{\boldsymbol{\xi}}\boldsymbol{\beta}'] \geq \int_U \boldsymbol{\beta}'^{\mathrm{T}}\mathbf{C}_{\boldsymbol{\xi}}^{\mathrm{T}}\mathbf{C}_{\boldsymbol{\xi}}\boldsymbol{\beta}' d\pi(\boldsymbol{\xi}) \geq \epsilon\pi(U) > 0.$$

This contradicts the hypothesis that $\boldsymbol{\beta}^{\mathrm{T}}\,\mathrm{E}[\mathbf{C}_{\boldsymbol{\xi}}^{\mathrm{T}}\mathbf{C}_{\boldsymbol{\xi}}]\boldsymbol{\beta} = 0$ for all $\boldsymbol{\beta} \in \mathbb{R}^{p+1}$. Hence the result follows by contradiction.

Proof of Theorem 3.2 *Part (i).* For $\boldsymbol{\theta} = (\boldsymbol{\beta}, \boldsymbol{\epsilon}) \in \mathbb{R}^{p+1} \times \mathcal{E}$ and $\rho \in S_n$, and given $\boldsymbol{\xi} \in \Xi$, we have that $R[\rho^{-1}(\boldsymbol{\theta}), \boldsymbol{\xi}] = R[\boldsymbol{\theta}, \rho(\boldsymbol{\xi})]$, since

$$\begin{aligned}
&R[\rho^{-1}(\boldsymbol{\theta}), \boldsymbol{\xi}] = \ell[\rho^{-1}(\boldsymbol{\theta}), h(\boldsymbol{\xi}, \boldsymbol{\gamma}_{\rho^{-1}(\boldsymbol{\theta}),\boldsymbol{\xi}})] \\
&= \left\|\boldsymbol{\Lambda}\boldsymbol{\beta} - h(\boldsymbol{\xi},\, \boldsymbol{\gamma}_{\rho^{-1}(\boldsymbol{\theta}),\boldsymbol{\xi}})\right\|^2 \\
&= \left\|\boldsymbol{\Lambda}\boldsymbol{\beta} - h(\rho(\boldsymbol{\xi}),\, \rho(\boldsymbol{\gamma}_{\rho^{-1}(\boldsymbol{\theta}),\boldsymbol{\xi}}))\right\|^2 \quad \text{by invariance of } \hat{\boldsymbol{\alpha}} \\
&= \left\|\boldsymbol{\Lambda}\boldsymbol{\beta} - h(\rho(\boldsymbol{\xi}),\, \boldsymbol{\gamma}_{\boldsymbol{\theta},\rho(\boldsymbol{\xi})})\right\|^2 \quad \text{by Lemma B.2} \\
&= R[\boldsymbol{\theta}, \rho(\boldsymbol{\xi})]\,.
\end{aligned}$$

Moreover, for the RDS π, we have that

$$R[\rho^{-1}(\boldsymbol{\theta}), \pi] = \int_\Xi R[\rho^{-1}(\boldsymbol{\theta}), \boldsymbol{\xi}]d\pi(\boldsymbol{\xi}) = \int_\Xi R[\boldsymbol{\theta}, \rho(\boldsymbol{\xi})]d\pi(\boldsymbol{\xi}) = R[\boldsymbol{\theta}, \rho_*\pi]\,.$$

Hence, for arbitrary $\boldsymbol{\theta} \in \Theta$,

$$\max_{\boldsymbol{\theta}' \in \Theta} R[\boldsymbol{\theta}', \pi] \geq \max_{\rho \in S_n} R[\rho^{-1}(\boldsymbol{\theta}), \pi] = \max_{\rho \in S_n} R[\boldsymbol{\theta}, \rho_*(\pi)] \geq \frac{1}{n!} \sum_{\rho \in S_n} R[\boldsymbol{\theta}, \rho_* \pi] = R[\boldsymbol{\theta}, \tilde{\pi}]\,,$$

and so $\Psi(\tilde{\pi}) = \max_{\boldsymbol{\theta}' \in \Theta} R[\boldsymbol{\theta}', \tilde{\pi}] \leq \max_{\boldsymbol{\theta}' \in \Theta} R[\boldsymbol{\theta}', \pi] = \Psi(\pi)$ as claimed.

Parts (ii) and (iii). Let the estimator be $\hat{\boldsymbol{\alpha}}_{\boldsymbol{\theta},\boldsymbol{\xi}} = \mathbf{A}_{\boldsymbol{\xi}} \mathbf{y} = \mathbf{A}_{\boldsymbol{\xi}} \boldsymbol{\gamma}_{\boldsymbol{\theta},\boldsymbol{\xi}}$. Suppose that $\boldsymbol{\xi} \sim \pi$ and $\rho \sim \text{Uniform}(S_n)$, so that $\rho(\boldsymbol{\xi}) \sim \tilde{\pi}$. Conditional on $\boldsymbol{\xi}$, the expected loss with design $\rho(\boldsymbol{\xi})$ is

$$\mathrm{E}[(\hat{\boldsymbol{\alpha}}_{\boldsymbol{\theta},\rho(\boldsymbol{\xi})} - \boldsymbol{\Lambda}\boldsymbol{\beta})^{\mathrm{T}} (\hat{\boldsymbol{\alpha}}_{\boldsymbol{\theta},\rho(\boldsymbol{\xi})} - \boldsymbol{\Lambda}\boldsymbol{\beta}) \,|\, \boldsymbol{\xi}] = \mathrm{cbias}(\hat{\boldsymbol{\alpha}}_{\boldsymbol{\theta},\rho(\boldsymbol{\xi})})^{\mathrm{T}}\, \mathrm{cbias}(\hat{\boldsymbol{\alpha}}_{\boldsymbol{\theta},\rho(\boldsymbol{\xi})}) + \mathrm{tr}(\mathrm{cvar}(\hat{\boldsymbol{\alpha}}_{\boldsymbol{\theta},\rho(\boldsymbol{\xi})}))\,,$$

where cbias and cvar denote the conditional bias and variance of $\hat{\boldsymbol{\alpha}}_{\boldsymbol{\theta},\rho(\boldsymbol{\xi})}$ given $\boldsymbol{\xi}$. Note that by invariance and Lemma B.2 $\hat{\boldsymbol{\alpha}}_{\boldsymbol{\theta},\rho(\boldsymbol{\xi})} = \hat{\boldsymbol{\alpha}}_{\rho^{-1}(\boldsymbol{\theta}),\boldsymbol{\xi}} = \mathbf{A}_{\boldsymbol{\xi}} \mathbf{F}_{\boldsymbol{\xi}} \boldsymbol{\beta} + \mathbf{A}_{\boldsymbol{\xi}} \rho^{-1}(\boldsymbol{\epsilon})(\boldsymbol{\xi})$. Moreover, by the argument from the proof of Proposition 3.1, given $\boldsymbol{\xi}$ the vector $\rho^{-1}(\boldsymbol{\epsilon})(\boldsymbol{\xi})$ has conditional expectation $\mathbf{0}$ and conditional variance matrix $\mathbf{V}(\boldsymbol{\xi})$. Hence $\mathrm{cbias}(\hat{\boldsymbol{\alpha}}_{\theta,\rho(\boldsymbol{\xi})}) = (\mathbf{A}_{\boldsymbol{\xi}} \mathbf{F}_{\boldsymbol{\xi}} - \boldsymbol{\Lambda})\boldsymbol{\beta}$ and $\mathrm{cvar}(\hat{\boldsymbol{\alpha}}_{\boldsymbol{\theta},\rho(\boldsymbol{\xi})}) = \mathbf{A}_{\boldsymbol{\xi}} \mathbf{V}(\boldsymbol{\xi}) \mathbf{A}_{\boldsymbol{\xi}}^{\mathrm{T}}$. Thus, by the law of total expectation, the marginal expected loss is

$$\begin{aligned} R[\boldsymbol{\theta}, \tilde{\pi}] &= \mathrm{E}[(\hat{\boldsymbol{\alpha}}_{\boldsymbol{\theta},\rho(\boldsymbol{\xi})} - \boldsymbol{\alpha})^{\mathrm{T}} (\hat{\boldsymbol{\alpha}}_{\boldsymbol{\theta},\rho(\boldsymbol{\xi})} - \boldsymbol{\alpha})] \\ &= \boldsymbol{\beta}^{\mathrm{T}}\, \mathrm{E}[(\mathbf{A}_{\boldsymbol{\xi}} \mathbf{F}_{\boldsymbol{\xi}} - \boldsymbol{\Lambda})^{\mathrm{T}} (\mathbf{A}_{\boldsymbol{\xi}} \mathbf{F}_{\boldsymbol{\xi}} - \boldsymbol{\Lambda})] \boldsymbol{\beta} + \mathrm{tr}\, \mathrm{E}[\mathbf{A}_{\boldsymbol{\xi}} \mathbf{V}(\boldsymbol{\xi}) \mathbf{A}_{\boldsymbol{\xi}}^{\mathrm{T}}]\,. \end{aligned}$$

To ensure that $\sup_{\boldsymbol{\theta} \in \Theta} R[\boldsymbol{\theta}, \tilde{\pi}] < \infty$ we require that $\mathrm{E}[\boldsymbol{\beta}^{\mathrm{T}} (\mathbf{A}_{\boldsymbol{\xi}} \mathbf{F}_{\boldsymbol{\xi}} - \boldsymbol{\Lambda})^{\mathrm{T}} (\mathbf{A}_{\boldsymbol{\xi}} \mathbf{F}_{\boldsymbol{\xi}} - \boldsymbol{\Lambda}) \boldsymbol{\beta}] = 0$ for all $\boldsymbol{\beta} \in \mathbb{R}^{p+1}$. If this condition does not hold, then the maximum expected loss is infinite and the estimator is not minimax. As $\mathbf{A}_{\boldsymbol{\xi}}$ is assumed continuous, by Lemma B.3 the above implies that $\mathbf{A}_{\boldsymbol{\xi}} \mathbf{F}_{\boldsymbol{\xi}} = \boldsymbol{\Lambda}$ for all $\boldsymbol{\xi} \in \mathrm{supp}(\pi)$.

Let $\preceq$ denote the Loewner inequality, i.e. $\mathbf{M}_1 \preceq \mathbf{M}_2$ if $\mathbf{M}_2 - \mathbf{M}_1$ is non-negative definite. If $\sup_{\boldsymbol{\epsilon} \in \mathcal{E}} S^2(\mathbf{x}) \equiv \sigma^2$ for all $\mathbf{x} \in \mathcal{X}$, then we have that

$$\begin{aligned} \mathbf{A}_{\boldsymbol{\xi}} \mathbf{V}(\boldsymbol{\xi}) \mathbf{A}_{\boldsymbol{\xi}}^{\mathrm{T}} &= \frac{n}{n-1} \mathbf{A}_{\boldsymbol{\xi}}\, \mathrm{diag}[S^2(\mathbf{x}_1), \ldots, S^2(\mathbf{x}_n)] \mathbf{A}_{\boldsymbol{\xi}}^{\mathrm{T}} - \frac{1}{n(n-1)} (\mathbf{E}(\boldsymbol{\xi}) \mathbf{A}_{\boldsymbol{\xi}}^{\mathrm{T}})^{\mathrm{T}} \mathbf{E}(\boldsymbol{\xi}) \mathbf{A}_{\boldsymbol{\xi}}^{\mathrm{T}} \\ &\preceq \frac{n\sigma^2}{n-1} \mathbf{A}_{\boldsymbol{\xi}} \mathbf{A}_{\boldsymbol{\xi}}^{\mathrm{T}}\,, \end{aligned}$$

and so

$$R[\boldsymbol{\theta}, \tilde{\pi}] = \operatorname{tr} \mathrm{E}[\mathbf{A}_{\boldsymbol{\xi}} \mathbf{V}(\boldsymbol{\xi}) \mathbf{A}_{\boldsymbol{\xi}}^{\mathrm{T}}] \leq \frac{n\sigma^2}{n-1} \mathrm{E}[\operatorname{tr} \mathbf{A}_{\boldsymbol{\xi}} \mathbf{A}_{\boldsymbol{\xi}}^{\mathrm{T}}].$$

Moreover, the Loewner bound is attained in the sense that $\mathbf{A}_{\boldsymbol{\xi}} \mathbf{V}(\boldsymbol{\xi}) \mathbf{A}_{\boldsymbol{\xi}}^{\mathrm{T}} = \frac{n\sigma^2}{n-1} \mathbf{A}_{\boldsymbol{\xi}} \mathbf{A}_{\boldsymbol{\xi}}^{\mathrm{T}}$ for all $\boldsymbol{\xi} \in \Xi$ when $\boldsymbol{\epsilon}(\mathbf{x}) = \sqrt{n}\sigma\tilde{\boldsymbol{\epsilon}}$ for all $\mathbf{x} \in \mathcal{X}$, with $\tilde{\boldsymbol{\epsilon}} \in \mathbb{R}^n$, $\sum_{i=1}^n \tilde{\epsilon}_i = 0$ and $\sum_{i=1}^n \tilde{\epsilon}_i^2 = 1$ (see next paragraph). This coincides with the case where the unit effects do not depend on the treatment, i.e. with unit-treatment additivity. Note that the above choice of $\boldsymbol{\epsilon}$ is an element of both $\mathcal{E}_1$ and $\mathcal{E}_2$. Thus, for $\Theta = \Theta_1$ or Θ_2,

$$\max_{\boldsymbol{\theta} \in \Theta} R(\boldsymbol{\theta}, \tilde{\pi}) = \max_{\boldsymbol{\epsilon} \in \mathcal{E}} \mathrm{E}(\operatorname{tr} \mathbf{A}_{\boldsymbol{\xi}} \mathbf{V}(\boldsymbol{\xi}) \mathbf{A}_{\boldsymbol{\xi}}^{\mathrm{T}}) = \frac{n\sigma^2}{n-1} \mathrm{E} \operatorname{tr}(\mathbf{A}_{\boldsymbol{\xi}} \mathbf{A}_{\boldsymbol{\xi}}^{\mathrm{T}}).$$

By the Gauss-Markov theorem, we have that if $\mathbf{A}_{\boldsymbol{\xi}} \mathbf{F}_{\boldsymbol{\xi}} = \boldsymbol{\Lambda}$ and $\mathbf{M}_{\boldsymbol{\xi}}$ is invertible, then $\mathbf{A}_{\boldsymbol{\xi}} \mathbf{A}_{\boldsymbol{\xi}}^{\mathrm{T}} \succeq \boldsymbol{\Lambda} \mathbf{M}_{\boldsymbol{\xi}}^{-1} \boldsymbol{\Lambda}^{\mathrm{T}}$ with equality if $\mathbf{A}_{\boldsymbol{\xi}} \mathbf{y}$ is equal to the ordinary least squares estimator, i.e. $\mathbf{A}_{\boldsymbol{\xi}} = \boldsymbol{\Lambda} \mathbf{M}_{\boldsymbol{\xi}}^{-1} \mathbf{F}_{\boldsymbol{\xi}}^{\mathrm{T}}$. Since trace respects the Loewner order,

$$\max_{\boldsymbol{\theta} \in \Theta} R(\boldsymbol{\theta}, \tilde{\pi}) \geq \frac{n\sigma^2}{n-1} \mathrm{E} \operatorname{tr}(\boldsymbol{\Lambda} \mathbf{M}_{\boldsymbol{\xi}}^{-1} \boldsymbol{\Lambda}^{\mathrm{T}}) = \frac{n\sigma^2}{n-1} \mathrm{E} \operatorname{tr}(\boldsymbol{\Lambda}^{\mathrm{T}} \boldsymbol{\Lambda} \mathbf{M}_{\boldsymbol{\xi}}^{-1}) \geq \frac{n\sigma^2}{n-1} \operatorname{tr}(\boldsymbol{\Lambda}^{\mathrm{T}} \boldsymbol{\Lambda} \mathbf{M}_{\boldsymbol{\xi}_L^*}^{-1}),$$

with equality if $\hat{\boldsymbol{\alpha}} = \hat{\boldsymbol{\alpha}}_{OLS}$ and π is a point mass measure on the L-optimal design $\boldsymbol{\xi}_L^*$. Hence in this case the minimax combination of estimator and RDS is ordinary least squares estimator together with complete randomization of a homoscedastic L-optimal design.

To verify that $\mathbf{A}_{\boldsymbol{\xi}} \mathbf{V}(\boldsymbol{\xi}) \mathbf{A}_{\boldsymbol{\xi}}^{\mathrm{T}} = \sigma^2 \mathbf{A}_{\boldsymbol{\xi}} \mathbf{A}_{\boldsymbol{\xi}}^{\mathrm{T}}$ when $\boldsymbol{\epsilon}(\mathbf{x}) = \sqrt{n}\sigma\tilde{\boldsymbol{\epsilon}}$, first note that $\mathbf{A}_{\boldsymbol{\xi}} \mathbf{1}_n = \mathbf{0}$ for all $\boldsymbol{\xi} \in \operatorname{supp}(\pi)$. This is true because $\mathbf{A}_{\boldsymbol{\xi}} \mathbf{F}_{\boldsymbol{\xi}} = \boldsymbol{\Lambda}$, and the first column of $\mathbf{A}_{\boldsymbol{\xi}} \mathbf{F}_{\boldsymbol{\xi}}$ equals $\mathbf{A}_{\boldsymbol{\xi}} \mathbf{1}_n$, while that of $\boldsymbol{\Lambda}$ equals $\mathbf{0}$. Thus, when $\boldsymbol{\epsilon}(\mathbf{x}) = \sigma\tilde{\boldsymbol{\epsilon}}$ we have that $\mathbf{E}(\boldsymbol{\xi}) = \sqrt{n}\sigma\tilde{\boldsymbol{\epsilon}} \mathbf{1}_n^{\mathrm{T}}$ and so $\mathbf{E}(\boldsymbol{\xi}) \mathbf{A}_{\boldsymbol{\xi}}^{\mathrm{T}} = \sqrt{n}\sigma\tilde{\boldsymbol{\epsilon}} \mathbf{1}_n^{\mathrm{T}} \mathbf{A}_{\boldsymbol{\xi}}^{\mathrm{T}} = \sqrt{n}\sigma\tilde{\boldsymbol{\epsilon}} (\mathbf{A}_{\boldsymbol{\xi}} \mathbf{1}_n)^{\mathrm{T}} = \mathbf{0}$. Moreover, we have that $S^2(\mathbf{x}) = \sigma^2$ for all $\mathbf{x}$. Hence in this case $\mathbf{A}_{\boldsymbol{\xi}} \mathbf{V}(\boldsymbol{\xi}) \mathbf{A}_{\boldsymbol{\xi}}^{\mathrm{T}} = \sigma^2 \mathbf{A}_{\boldsymbol{\xi}} \mathbf{A}_{\boldsymbol{\xi}}^{\mathrm{T}}$.

Lemma B.4. *λ is a non-zero eigenvalue of $\mathbf{Q}_{\boldsymbol{\xi}} = \mathbf{F}_{\boldsymbol{\xi}} \mathbf{M}_{\boldsymbol{\xi}}^{-1} \boldsymbol{\Lambda}^{\mathrm{T}} \boldsymbol{\Lambda} \mathbf{M}_{\boldsymbol{\xi}}^{-1} \mathbf{F}_{\boldsymbol{\xi}}^{\mathrm{T}}$ if and only if it is a non-zero eigenvalue of $\boldsymbol{\Lambda}^{\mathrm{T}} \boldsymbol{\Lambda} \mathbf{M}_{\boldsymbol{\xi}}^{-1}$.*

Proof. We first show that if λ is a non-zero eigenvalue of $\mathbf{Q}_{\boldsymbol{\xi}}$ then it is an eigenvalue of $\boldsymbol{\Lambda}^{\mathrm{T}} \boldsymbol{\Lambda} \mathbf{M}_{\boldsymbol{\xi}}^{-1}$. Let $\mathbf{v} \in \mathbb{R}^n$ be a corresponding eigenvector, i.e. $\mathbf{Q}_{\boldsymbol{\xi}} \mathbf{v} = \lambda \mathbf{v}$. Then due to the form of $\mathbf{Q}_{\boldsymbol{\xi}}$ we have $\mathbf{v} = \mathbf{F}_{\boldsymbol{\xi}}\mathbf{w}$ for some $\mathbf{w} \in \mathbb{R}^{p+1}$. After some algebra, we see that $\lambda\mathbf{M}_{\boldsymbol{\xi}}\mathbf{w} = \boldsymbol{\Lambda}^{\mathrm{T}}\boldsymbol{\Lambda}\mathbf{w}$. Hence $\mathbf{M}_{\boldsymbol{\xi}}\mathbf{w}$ is an eigenvector of $\boldsymbol{\Lambda}^{\mathrm{T}}\boldsymbol{\Lambda}\mathbf{M}_{\boldsymbol{\xi}}^{-1}$ with eigenvalue λ as required.

Now we show that if λ is an eigenvalue of $\boldsymbol{\Lambda}^{\mathrm{T}}\boldsymbol{\Lambda}\mathbf{M}_{\boldsymbol{\xi}}^{-1}$ then it is an eigenvalue of $\mathbf{Q}_{\boldsymbol{\xi}}$. Let $\mathbf{z} \in \mathbb{R}^{p+1}$ be such that $\boldsymbol{\Lambda}^{\mathrm{T}}\boldsymbol{\Lambda}\mathbf{M}_{\boldsymbol{\xi}}^{-1}\mathbf{z} = \lambda\mathbf{z}$. Then

$$\lambda\mathbf{F}_{\boldsymbol{\xi}}\mathbf{M}_{\boldsymbol{\xi}}^{-1}\mathbf{z} = \mathbf{F}_{\boldsymbol{\xi}}\mathbf{M}_{\boldsymbol{\xi}}^{-1}\boldsymbol{\Lambda}^{\mathrm{T}}\boldsymbol{\Lambda}\mathbf{M}_{\boldsymbol{\xi}}^{-1}\mathbf{z}\,. \tag{B.2}$$

Moreover, since $\mathbf{M}_{\boldsymbol{\xi}}$ is invertible and $\mathrm{rank}(\mathbf{F}_{\boldsymbol{\xi}}^{\mathrm{T}}) = \mathrm{rank}(\mathbf{M}_{\boldsymbol{\xi}})$, we must have that $\mathbf{F}_{\boldsymbol{\xi}}^{\mathrm{T}}$ is of full rank and so $\mathbf{z} = \mathbf{F}_{\boldsymbol{\xi}}^{\mathrm{T}}\mathbf{w}$ for some $\mathbf{w} \in \mathbb{R}^{n}$. Combining this with (B.2), we see that $\lambda\mathbf{H}_{\boldsymbol{\xi}}\mathbf{w} = \mathbf{Q}_{\boldsymbol{\xi}}\mathbf{w}$, with $\mathbf{H}_{\boldsymbol{\xi}}$ the hat matrix. Some further algebra shows that $\mathbf{Q}_{\boldsymbol{\xi}} = \mathbf{Q}_{\boldsymbol{\xi}}\mathbf{H}_{\boldsymbol{\xi}}$. Hence $\mathbf{H}_{\boldsymbol{\xi}}\mathbf{w}$ is an eigenvector of $\mathbf{Q}_{\boldsymbol{\xi}}$ with eigenvalue λ. $\square$

Proof of Proposition 3.3. *Part (i).* Assume that $\mathbf{F}_{\boldsymbol{\xi}}^{\mathrm{T}}\mathbf{F}_{\boldsymbol{\xi}}$ is invertible. Recall that the response is $\mathbf{y} = \mathbf{F}_{\boldsymbol{\xi}}\boldsymbol{\beta} + \boldsymbol{\epsilon}(\boldsymbol{\xi})$ for some $\boldsymbol{\epsilon} \in \mathcal{E}$. After some algebra we find that $\boldsymbol{\alpha} - \hat{\boldsymbol{\alpha}}_{\mathrm{OLS}} = -\boldsymbol{\Lambda}(\mathbf{F}_{\boldsymbol{\xi}}^{\mathrm{T}}\mathbf{F}_{\boldsymbol{\xi}})^{-1}\mathbf{F}_{\boldsymbol{\xi}}^{\mathrm{T}}\boldsymbol{\epsilon}(\boldsymbol{\xi})$ for some $\boldsymbol{\epsilon} \in \mathcal{E}$. Hence, given $\mathbf{y} \in \mathbb{R}^{n}$, the set, $\mathcal{A}$, of feasible values for $\boldsymbol{\alpha}$ is a compact set centred on $\hat{\boldsymbol{\alpha}}_{\mathrm{OLS}}$. Moreover, since $\mathcal{E}$ is symmetric (i.e. invariant under multiplication by -1), the set $\mathcal{A}$ is symmetric in the sense that $\mathbf{a} = \hat{\boldsymbol{\alpha}}_{\mathrm{OLS}} + \mathbf{d} \in \mathcal{A}$ if and only if $\mathbf{a}^{-} = \hat{\boldsymbol{\alpha}}_{OLS} - \mathbf{d} \in \mathcal{A}$. We argue below that this symmetry means that the choice for $\hat{\boldsymbol{\alpha}}$ giving minimax loss is $\hat{\boldsymbol{\alpha}}_{\mathrm{OLS}}$.

From the above, the set of feasible values for the vector $\boldsymbol{\alpha} - \hat{\boldsymbol{\alpha}}_{\mathrm{OLS}}$ is

$$\mathcal{D}_{\boldsymbol{\xi}} = \{-\boldsymbol{\Lambda}(\mathbf{F}_{\boldsymbol{\xi}}^{\mathrm{T}}\mathbf{F}_{\boldsymbol{\xi}})^{-1}\mathbf{F}_{\boldsymbol{\xi}}^{\mathrm{T}}\boldsymbol{\epsilon}(\boldsymbol{\xi}) \mid \boldsymbol{\epsilon} \in \mathcal{E}\}\,.$$

By symmetry of $\mathcal{E}$, if $\mathbf{d} \in \mathcal{D}_{\boldsymbol{\xi}}$ then $-\mathbf{d} \in \mathcal{D}_{\boldsymbol{\xi}}$. Now let $\hat{\boldsymbol{\alpha}}$ be any other estimate of $\boldsymbol{\alpha}$. We have that $\ell(\boldsymbol{\theta}, \hat{\boldsymbol{\alpha}}) = \|\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}\|^2 = \|\hat{\boldsymbol{\alpha}} - \hat{\boldsymbol{\alpha}}_{\mathrm{OLS}}\|^2 + 2\langle\hat{\boldsymbol{\alpha}} - \hat{\boldsymbol{\alpha}}_{\mathrm{OLS}}, \hat{\boldsymbol{\alpha}}_{\mathrm{OLS}} - \boldsymbol{\alpha}\rangle + \|\hat{\boldsymbol{\alpha}}_{\mathrm{OLS}} - \boldsymbol{\alpha}\|^2 \geq 2\langle\hat{\boldsymbol{\alpha}} - \hat{\boldsymbol{\alpha}}_{\mathrm{OLS}}, \hat{\boldsymbol{\alpha}}_{\mathrm{OLS}} - \boldsymbol{\alpha}\rangle + \|\hat{\boldsymbol{\alpha}}_{\mathrm{OLS}} - \boldsymbol{\alpha}\|^2$. Hence

$$\max_{\boldsymbol{\epsilon}\in\mathcal{E}} \|\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}\|^2 \geq \max_{\mathbf{d}\in\mathcal{D}_{\boldsymbol{\xi}}}\{2\langle\hat{\boldsymbol{\alpha}} - \hat{\boldsymbol{\alpha}}_{\mathrm{OLS}}, \mathbf{d}\rangle + \|\mathbf{d}\|^2\}. \tag{B.3}$$

We claim that the right hand side of (B.3) is at least $\max_{\mathbf{d}\in\mathcal{D}_{\boldsymbol{\xi}}} \|\mathbf{d}\|^2 = \max_{\boldsymbol{\epsilon}\in\mathcal{E}} \|\hat{\boldsymbol{\alpha}}_{\mathrm{OLS}} - \boldsymbol{\alpha}\|^2$. To see this, suppose that $\mathbf{d}^* \in \mathcal{D}_{\boldsymbol{\xi}}$ maximizes $\|\mathbf{d}\|^2$. If $\langle\hat{\boldsymbol{\alpha}} - \hat{\boldsymbol{\alpha}}_{\mathrm{OLS}}, \mathbf{d}^*\rangle \geq 0$, then $\max_{\mathbf{d}\in\mathcal{D}_{\boldsymbol{\xi}}}\{2\langle\hat{\boldsymbol{\alpha}}-$

$\hat{\boldsymbol{\alpha}}_{\mathrm{OLS}}, \mathbf{d}\rangle + \|\mathbf{d}\|^2\} \geq 2\langle\hat{\boldsymbol{\alpha}} - \hat{\boldsymbol{\alpha}}_{\mathrm{OLS}}, \mathbf{d}^*\rangle + \|\mathbf{d}^*\|^2 \geq \max_{\mathbf{d}\in\mathcal{D}_{\boldsymbol{\xi}}} \|\mathbf{d}\|^2$ as claimed. However, if $\langle\hat{\boldsymbol{\alpha}} - \hat{\boldsymbol{\alpha}}_{\mathrm{OLS}}, \mathbf{d}^*\rangle < 0$ then considering $-\mathbf{d}^* \in \mathcal{D}_{\boldsymbol{\xi}}$ we see that $\max_{\mathbf{d}\in\mathcal{D}_{\boldsymbol{\xi}}}\{2\langle\hat{\boldsymbol{\alpha}} - \hat{\boldsymbol{\alpha}}_{\mathrm{OLS}}, \mathbf{d}\rangle + \|\mathbf{d}\|^2\} \geq 2\langle\hat{\boldsymbol{\alpha}} - \hat{\boldsymbol{\alpha}}_{\mathrm{OLS}}, -\mathbf{d}^*\rangle + \|-\mathbf{d}^*\|^2 > \|\mathbf{d}^*\|^2 = \max_{\mathbf{d}\in\mathcal{D}_{\boldsymbol{\xi}}} \|\mathbf{d}\|^2$. Thus in either case $\max_{\boldsymbol{\epsilon}\in\mathcal{E}} \|\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}\|^2 \geq \max_{\boldsymbol{\epsilon}\in\mathcal{E}} \|\hat{\boldsymbol{\alpha}}_{\mathrm{OLS}} - \boldsymbol{\alpha}\|^2$, and so $\hat{\boldsymbol{\alpha}}_{\mathrm{OLS}}$ minimizes the maximum loss.

Part (ii). Consider the linear estimator $\hat{\boldsymbol{\alpha}} = \mathbf{A}_{\boldsymbol{\xi}}\mathbf{y} = \mathbf{A}_{\boldsymbol{\xi}}\boldsymbol{\gamma}_{\boldsymbol{\theta},\boldsymbol{\xi}}$. As $\boldsymbol{\xi}$ and $\boldsymbol{\gamma}_{\boldsymbol{\theta},\boldsymbol{\xi}}$ are deterministic we have that $R(\boldsymbol{\theta};\boldsymbol{\xi}) = \ell(\boldsymbol{\theta}, \mathbf{A}_{\boldsymbol{\xi}}\boldsymbol{\gamma}_{\boldsymbol{\theta},\boldsymbol{\xi}})$. The loss is

$$\ell(\boldsymbol{\theta}, \mathbf{A}_{\boldsymbol{\xi}}\boldsymbol{\gamma}_{\boldsymbol{\theta},\boldsymbol{\xi}}) = \boldsymbol{\beta}^{\mathrm{T}}(\mathbf{A}_{\boldsymbol{\xi}}\mathbf{F}_{\boldsymbol{\xi}} - \boldsymbol{\Lambda})^{\mathrm{T}}(\mathbf{A}_{\boldsymbol{\xi}}\mathbf{F}_{\boldsymbol{\xi}} - \boldsymbol{\Lambda})\boldsymbol{\beta} + 2\boldsymbol{\beta}^{\mathrm{T}}(\mathbf{A}_{\boldsymbol{\xi}}\mathbf{F}_{\boldsymbol{\xi}} - \boldsymbol{\Lambda})^{\mathrm{T}}\mathbf{A}_{\boldsymbol{\xi}}\boldsymbol{\epsilon}(\boldsymbol{\xi}) + \boldsymbol{\epsilon}(\boldsymbol{\xi})^{\mathrm{T}}\mathbf{A}_{\boldsymbol{\xi}}^{\mathrm{T}}\mathbf{A}_{\boldsymbol{\xi}}\boldsymbol{\epsilon}(\boldsymbol{\xi}),$$

For the OLS estimator we have $\mathbf{A}_{\boldsymbol{\xi}}\mathbf{F}_{\boldsymbol{\xi}} = \boldsymbol{\Lambda}\mathbf{M}_{\boldsymbol{\xi}}^{-1}\mathbf{F}_{\boldsymbol{\xi}}^{\mathrm{T}}\mathbf{F}_{\boldsymbol{\xi}} = \boldsymbol{\Lambda}$. Therefore the loss simplifies to $\ell(\boldsymbol{\theta}, \hat{\boldsymbol{\alpha}}) = \boldsymbol{\epsilon}(\boldsymbol{\xi})^{\mathrm{T}}\mathbf{Q}_{\boldsymbol{\xi}}\boldsymbol{\epsilon}(\boldsymbol{\xi})$, with $\mathbf{Q}_{\boldsymbol{\xi}} = \mathbf{F}_{\boldsymbol{\xi}}\mathbf{M}_{\boldsymbol{\xi}}^{-1}\boldsymbol{\Lambda}^{\mathrm{T}}\boldsymbol{\Lambda}\mathbf{M}_{\boldsymbol{\xi}}^{-1}\mathbf{F}_{\boldsymbol{\xi}}^{\mathrm{T}}$ and so

$$\max_{\boldsymbol{\theta}\in\Theta_2} R(\boldsymbol{\theta};\boldsymbol{\xi}) = \max_{\mathbf{e}\in\mathbb{R}^n:\mathbf{e}^{\mathrm{T}}\mathbf{1}=0,\mathbf{e}^{\mathrm{T}}\mathbf{e}\leq n\sigma^2} \mathbf{e}^{\mathrm{T}}\mathbf{Q}_{\boldsymbol{\xi}}\mathbf{e} \leq n\sigma^2\lambda_{\max}(\mathbf{Q}_{\boldsymbol{\xi}}).$$

The upper bound on the right hand side is attained when $\mathbf{e} = \sqrt{n}\sigma\mathbf{v}$, with $\mathbf{v}$ a principal normalized eigenvector of $\mathbf{Q}_{\boldsymbol{\xi}}$. Moreover, we claim that $\mathbf{e}^{\mathrm{T}}\mathbf{1} = 0$ (see next paragraph) and so $\boldsymbol{\epsilon}(\mathbf{x}) \equiv \mathbf{e} = \sqrt{n}\sigma\mathbf{v} \in \mathcal{E}_2$. Hence

$$\max_{\boldsymbol{\theta}\in\Theta_2} R(\boldsymbol{\theta};\boldsymbol{\xi}) = n\sigma^2\lambda_{\max}(\mathbf{Q}_{\boldsymbol{\xi}}) = n\sigma^2\lambda_{\max}(\boldsymbol{\Lambda}^{\mathrm{T}}\boldsymbol{\Lambda}\mathbf{M}_{\boldsymbol{\xi}}^{-1}),$$

where the last equality follows by Lemma B.4. This establishes the result.

To see that $\mathbf{e}^{\mathrm{T}}\mathbf{1} = 0$, note that $\mathbf{1} = \mathbf{F}_{\boldsymbol{\xi}}[1, 0, \ldots, 0]^{\mathrm{T}}$ due to the intercept term in the model. Hence

$$\mathbf{Q}_{\boldsymbol{\xi}}\mathbf{1} = \mathbf{F}_{\boldsymbol{\xi}}\mathbf{M}_{\boldsymbol{\xi}}^{-1}\boldsymbol{\Lambda}^{\mathrm{T}}\boldsymbol{\Lambda}\mathbf{M}_{\boldsymbol{\xi}}^{-1}\mathbf{F}_{\boldsymbol{\xi}}^{\mathrm{T}}\mathbf{F}_{\boldsymbol{\xi}}\begin{pmatrix}1\\0\\\vdots\\0\end{pmatrix} = \mathbf{F}_{\boldsymbol{\xi}}\mathbf{M}_{\boldsymbol{\xi}}^{-1}\boldsymbol{\Lambda}^{\mathrm{T}}\boldsymbol{\Lambda}\begin{pmatrix}1\\0\\\vdots\\0\end{pmatrix} = \mathbf{0}$$

and $\mathbf{1}$ is an eigenvector of $\mathbf{Q}_{\boldsymbol{\xi}}$ with eigenvalue zero. $\mathbf{Q}_{\boldsymbol{\xi}}$ is real and symmetric, so eigenvectors corresponding to distinct eigenvectors of $\mathbf{Q}_{\boldsymbol{\xi}}$ are orthogonal, and so $\mathbf{e}^{\mathrm{T}}\mathbf{1} = \sqrt{n}\sigma\mathbf{v}^{\mathrm{T}}\mathbf{1} = 0$.

Part (iii). We divide the proof into two cases. First we show that if $u \geq \Psi(\boldsymbol{\xi}; \Theta_2)$ then $\max_{\boldsymbol{\theta}' \in \Theta_2} S(\boldsymbol{\theta}', \boldsymbol{\xi}, u) = 0$. Note that, for a deterministic design, with $\boldsymbol{\theta} \in \Theta_2$ the loss $\ell(\boldsymbol{\theta}, \hat{\boldsymbol{\alpha}}) = R(\boldsymbol{\theta}, \boldsymbol{\xi}) = \mathbf{e}^{\mathrm{T}}\mathbf{Q}_{\boldsymbol{\xi}}\mathbf{e}$ is also deterministic, and bounded above via $\ell(\boldsymbol{\theta}, \hat{\boldsymbol{\alpha}}) \leq \Psi(\boldsymbol{\xi}; \Theta_2)$. Hence, if $u \geq \Psi(\boldsymbol{\xi}; \Theta_2)$, then $S(\boldsymbol{\theta}, \boldsymbol{\xi}, u) = \Pr[\ell(\boldsymbol{\theta}, \hat{\boldsymbol{\alpha}}) > u] = 0$ for all $\boldsymbol{\theta}$, and so $\max_{\boldsymbol{\theta}' \in \Theta_2} S(\boldsymbol{\theta}', \boldsymbol{\xi}, u) = 0$.

Next we show that if $u < \Psi(\boldsymbol{\xi}; \Theta_2)$ then $\max_{\boldsymbol{\theta}' \in \Theta_2} S(\boldsymbol{\theta}', \boldsymbol{\xi}, u) = 1$. Note that $S(\boldsymbol{\theta}', \boldsymbol{\xi}, u) \leq 1$ for all $\boldsymbol{\theta}'$, as it is a probability. By the proof of the previous part, $R(\boldsymbol{\theta}^*, \boldsymbol{\xi}) = \Psi(\boldsymbol{\xi}; \Theta_2)$ for some $\boldsymbol{\theta}^* \in \Theta_2$. Hence, if $u < \Psi(\boldsymbol{\xi}; \Theta_2)$ then $S(\boldsymbol{\theta}^*, \boldsymbol{\xi}, u) = \Pr[R(\boldsymbol{\theta}^*, \boldsymbol{\xi}) > u] = 1$. Hence $\max_{\boldsymbol{\theta}' \in \Theta_2} S(\boldsymbol{\theta}', \boldsymbol{\xi}, u) = 1$.

B.1.2 Numerical results for the example of Section 3.2

Table 1 gives details of (i) an L-optimal design, complete randomization of which gives a minimax RDS; and (ii) a poor choice of treatments, for which complete randomization gives a less efficient strategy than use of the unrandomized version of $\boldsymbol{\xi}_L^*$. The first design was computed using a multistart co-ordinate exchange algorithm.

B.2 Proofs and additional results for Section 4

Proposition B.5. *With loss function (6) and $\Theta = \mathcal{X} \times [\underline{\sigma}^2, \bar{\sigma}^2] \times \mathbb{R}^{p+1}$, and any non-singular RDS π, the least squares prediction $\hat{\mu}(\mathbf{x}) = \mathbf{f}^{\mathrm{T}}(\mathbf{x})\hat{\boldsymbol{\beta}}_{OLS}$ is minimax among linear estimators of the form $\mathbf{a}_{\boldsymbol{\xi},\mathbf{x}}^{\mathrm{T}}\mathbf{y}$ with $(\boldsymbol{\xi}', \mathbf{x}') \mapsto \mathbf{a}_{\boldsymbol{\xi}',\mathbf{x}'}$ a continuous map on $\mathrm{supp}(\pi) \times \mathcal{X}$.*

Proof of Proposition B.5 The argument is similar to that in the proof of Theorem 3.2. The expected loss with estimator $\mathbf{a}_{\boldsymbol{\xi},\mathbf{x}}^{\mathrm{T}}\mathbf{y}$ is

$$\begin{aligned}
\mathrm{MSE}[\mathbf{f}^{\mathrm{T}}(\mathbf{x})\hat{\boldsymbol{\beta}}] &= \mathrm{E}[\mathrm{E}[(\mathbf{a}_{\boldsymbol{\xi},\mathbf{x}}^{\mathrm{T}}\mathbf{y} - \mathbf{f}^{\mathrm{T}}(\mathbf{x})\boldsymbol{\beta})^2] \,|\, \boldsymbol{\xi}] = \mathrm{E}[\mathrm{cbias}_{\boldsymbol{\xi}}(\mathbf{a}_{\boldsymbol{\xi},\mathbf{x}}^{\mathrm{T}}\mathbf{y})^2 + \mathrm{cvar}_{\boldsymbol{\xi}}(\mathbf{a}_{\boldsymbol{\xi},\mathbf{x}}^{\mathrm{T}}\mathbf{y})]\,, \\
&= \boldsymbol{\beta}^{\mathrm{T}}\,\mathrm{E}[(\mathbf{a}_{\boldsymbol{\xi},\mathbf{x}}^{\mathrm{T}}\mathbf{F}_{\boldsymbol{\xi}} - \mathbf{f}^{\mathrm{T}}(\mathbf{x}))^{\mathrm{T}}(\mathbf{a}_{\boldsymbol{\xi},\mathbf{x}}^{\mathrm{T}}\mathbf{F}_{\boldsymbol{\xi}} - \mathbf{f}^{\mathrm{T}}(\mathbf{x}))]\boldsymbol{\beta} + \sigma^2\,\mathrm{E}[\mathbf{a}_{\boldsymbol{\xi},\mathbf{x}}^{\mathrm{T}}\mathbf{a}_{\boldsymbol{\xi},\mathbf{x}}]\,,
\end{aligned}$$

where $\mathrm{cbias}_{\boldsymbol{\xi}}$ and $\mathrm{cvar}_{\boldsymbol{\xi}}$ denote the conditional bias and variance given $\boldsymbol{\xi}$.

As $\boldsymbol{\beta}$ can take any value in $\mathbb{R}^{p+1}$, for the maximum expected loss to be bounded we require that $\boldsymbol{\beta}^{\mathrm{T}}\,\mathrm{E}[(\mathbf{a}_{\boldsymbol{\xi},\mathbf{x}}^{\mathrm{T}}\mathbf{F}_{\boldsymbol{\xi}} - \mathbf{f}^{\mathrm{T}}(\mathbf{x}))^{\mathrm{T}}(\mathbf{a}_{\boldsymbol{\xi},\mathbf{x}}^{\mathrm{T}}\mathbf{F}_{\boldsymbol{\xi}} - \mathbf{f}^{\mathrm{T}}(\mathbf{x}))]\boldsymbol{\beta} = 0$ for all $\boldsymbol{\beta} \in \mathbb{R}^{p+1}$, which implies that $\mathbf{a}_{\boldsymbol{\xi},\mathbf{x}}^{\mathrm{T}}\mathbf{F}_{\boldsymbol{\xi}} = \mathbf{f}^{\mathrm{T}}(\mathbf{x})$ for all $(\boldsymbol{\xi},\mathbf{x}) \in \mathrm{supp}(\pi) \times \mathcal{X}$ by Lemma B.3. The expected loss then simplifies to

$$\mathrm{MSE}[\mathbf{f}^{\mathrm{T}}(\mathbf{x})\hat{\boldsymbol{\beta}}] = \sigma^2\,\mathrm{E}[\mathbf{a}_{\boldsymbol{\xi},\mathbf{x}}^{\mathrm{T}}\mathbf{a}_{\boldsymbol{\xi},\mathbf{x}}]\,.$$

For all $\boldsymbol{\xi}' \in \mathrm{supp}(\pi)$, since $\mathbf{a}_{\boldsymbol{\xi}',\mathbf{x}}^{\mathrm{T}}\mathbf{F}_{\boldsymbol{\xi}'} = \mathbf{f}^{\mathrm{T}}(\mathbf{x})$ and $\mathbf{M}_{\boldsymbol{\xi}'}^{-1}$ exists by non-singularity of π, by the Gauss-Markov theorem we have $\mathbf{a}_{\boldsymbol{\xi}',\mathbf{x}}^{\mathrm{T}}\mathbf{a}_{\boldsymbol{\xi}',\mathbf{x}} \geq \mathbf{f}(\mathbf{x})^{\mathrm{T}}\mathbf{M}_{\boldsymbol{\xi}'}^{-1}\mathbf{f}(\mathbf{x})$. Hence for all $\boldsymbol{\theta}$ we have $R[\boldsymbol{\theta},\pi] = \mathrm{MSE}[\mathbf{a}_{\boldsymbol{\xi},\mathbf{x}}^{\mathrm{T}}\mathbf{y}] \geq \sigma^2\,\mathrm{E}[\mathbf{f}(\mathbf{x})^{\mathrm{T}}\mathbf{M}_{\boldsymbol{\xi}}^{-1}\mathbf{f}(\mathbf{x})]$, and also $\Psi(\pi) \geq \bar{\sigma}^2 \max_{\mathbf{x}'\in\mathcal{X}} \mathrm{E}[\mathbf{f}(\mathbf{x}')^{\mathrm{T}}\mathbf{M}_{\boldsymbol{\xi}}^{-1}\mathbf{f}(\mathbf{x}')]$. Moreover this lower bound is attained by choosing $\mathbf{a}_{\boldsymbol{\xi},\mathbf{x}}^{\mathrm{T}} = \mathbf{f}^{\mathrm{T}}(\mathbf{x})\mathbf{M}_{\boldsymbol{\xi}}^{-1}\mathbf{F}_{\boldsymbol{\xi}}^{\mathrm{T}}$, which is continuous in $(\boldsymbol{\xi},\mathbf{x})$ by the non-singularity of π. Hence the OLS prediction is minimax.

Proof of Proposition 4.1 Recall that $\ell(\boldsymbol{\theta},\hat{\boldsymbol{\alpha}}) = [\mathbf{f}^{\mathrm{T}}(\mathbf{x})\hat{\boldsymbol{\beta}} - \mathbf{f}^{\mathrm{T}}(\mathbf{x})\boldsymbol{\beta}]^2$, where $\hat{\boldsymbol{\beta}} = (\mathbf{F}_{\boldsymbol{\xi}}^{\mathrm{T}}\mathbf{F}_{\boldsymbol{\xi}})^{-1}\mathbf{F}_{\boldsymbol{\xi}}^{\mathrm{T}}\mathbf{y}$. We have that $\mathbf{f}^{\mathrm{T}}(\mathbf{x})\hat{\boldsymbol{\beta}} - \mathbf{f}^{\mathrm{T}}(\mathbf{x})\boldsymbol{\beta} \,|\, \boldsymbol{\xi},\sigma^2 \sim N[0, \sigma^2\mathbf{f}^{\mathrm{T}}(\mathbf{x})(\mathbf{F}_{\boldsymbol{\xi}}^{\mathrm{T}}\mathbf{F}_{\boldsymbol{\xi}})^{-1}\mathbf{f}(\mathbf{x})]$ and so

$$\begin{aligned}
&\Pr[\ell(\boldsymbol{\theta},\hat{\boldsymbol{\alpha}}) > u \,|\, \boldsymbol{\xi},\sigma^2] = 1 - \Pr\left[-\sqrt{u} \leq \mathbf{f}^{\mathrm{T}}(\mathbf{x})\hat{\boldsymbol{\beta}} - \mathbf{f}^{\mathrm{T}}(\mathbf{x})\boldsymbol{\beta} \leq \sqrt{u}\right] \\
&= 2 - 2\Phi\left[\frac{\sqrt{u}}{\sigma}\{\mathbf{f}^{\mathrm{T}}(\mathbf{x})(\mathbf{F}_{\boldsymbol{\xi}}^{\mathrm{T}}\mathbf{F}_{\boldsymbol{\xi}})^{-1}\mathbf{f}(\mathbf{x})\}^{-1/2}\right].
\end{aligned}$$

Hence, if $\boldsymbol{\xi} \sim \pi$ then prior to sampling $\boldsymbol{\xi}$ the survivor function satisfies

$$\begin{aligned}
S(\boldsymbol{\theta},\pi,u) &= \Pr[\ell(\boldsymbol{\theta},\hat{\boldsymbol{\alpha}}) > u] = \sum_{\boldsymbol{\xi}\in\mathrm{supp}(\pi)} \Pr[\ell(\boldsymbol{\theta},\hat{\boldsymbol{\alpha}}) > u \,|\, \boldsymbol{\xi}]\pi(\boldsymbol{\xi}) \\
&= \sum_{\boldsymbol{\xi}\in\mathrm{supp}(\pi)} \left(2 - 2\Phi\left[\frac{\sqrt{u}}{\sigma}\{\mathbf{f}^{\mathrm{T}}(\mathbf{x})(\mathbf{F}_{\boldsymbol{\xi}}^{\mathrm{T}}\mathbf{F}_{\boldsymbol{\xi}})^{-1}\mathbf{f}(\mathbf{x})\}^{-1/2}\right]\right)\pi(\boldsymbol{\xi})
\end{aligned}$$

Taking the maximum of this expression with respect to $\boldsymbol{\theta} = (\mathbf{x},\sigma^2,\boldsymbol{\beta}) \in \Theta = \mathcal{X} \times [\underline{\sigma}^2, \bar{\sigma}^2] \times B$ gives the result.

Theorem B.6 (Jensen's inequality, Ferguson 1967, p.76)**.** *Let g be a convex function defined in a convex subset C of n-dimensional Euclidean space R^n and let $\mathbf{X} = (X_1,\ldots,X_n)$ be an integrable random vector such that $P(\mathbf{X} \in C) = 1$. Then $\mathrm{E}(\mathbf{X}) \in C$, $\mathrm{E}[g(\mathbf{X})]$ exists, and*

$$g[\mathrm{E}(\mathbf{X})] \leq \mathrm{E}[g(\mathbf{X})]\,.$$

Lemma B.7. *If the RDS is non-singular, then* $\mathbf{M}_{\boldsymbol{\xi}}$ *and* $\mathbf{M}_{\boldsymbol{\xi}}^{-1}$ *are integrable random matrices, so* $\mathrm{E}[\mathbf{M}_{\boldsymbol{\xi}}]$ *and* $\mathrm{E}[\mathbf{M}_{\boldsymbol{\xi}}^{-1}]$ *exist.*

Proof. By Lemma B.1 the mappings $\mathbf{d} \mapsto \mathbf{M}_{\mathbf{d}}$ and $\mathbf{d} \mapsto \mathbf{M}_{\mathbf{d}}^{-1}$ defined on $\mathrm{supp}(\pi)$ are continuous, and so measurable. The continuous image of a compact set is compact, so the images of these mappings are bounded. Hence $\mathbf{M}_{\boldsymbol{\xi}}$ and $\mathbf{M}_{\boldsymbol{\xi}}^{-1}$ are bounded and so integrable. □

Lemma B.8. *Let* η^* *denote a G-optimal approximate design. Then:*

$$\max_{\mathbf{x}\in\mathcal{X}} \tilde{R}(\mathbf{x},\pi) = \max_{\mathbf{x}\in\mathcal{X}} \mathbf{f}^{\mathrm{T}}(\mathbf{x})\,\mathrm{E}[\mathbf{M}_{\boldsymbol{\xi}}^{-1}]\mathbf{f}(\mathbf{x}) \geq \frac{1}{n}\max_{\mathbf{x}\in\mathcal{X}} \mathbf{f}^{\mathrm{T}}(\mathbf{x})\mathbf{M}_{\eta^*}^{-1}\mathbf{f}(\mathbf{x})$$

Hence if there exists an RDS π *with* $\mathrm{E}[\mathbf{M}_{\boldsymbol{\xi}}^{-1}] = \frac{1}{n}\mathbf{M}_{\eta^*}^{-1}$ *then this is minimax.*

Proof. Let C be the set of symmetric positive definite matrices of the form $\sum_{k=1}^{K} a_k \mathbf{f}(\check{\mathbf{x}}_k)\mathbf{f}^{\mathrm{T}}(\check{\mathbf{x}}_k)$, for some $K \in \mathbb{N}$, $\check{\mathbf{x}}_1,\ldots,\check{\mathbf{x}}_K \in \mathcal{X}$, and positive reals a_k with $\sum_{k=1}^{K} a_k = n$. Note that this is a convex subset of $\mathbb{R}^{(p+1)^2}$, and that for any $\mathbf{A} \in C$ we have $\mathbf{A} = n\mathbf{M}_\eta$ for some approximate design η.

Given any $\mathbf{x} \in \mathcal{X}$, the function $g_{\mathbf{x}} : C \to \mathbb{R}$ defined by $g_{\mathbf{x}}(\mathbf{A}) = \mathbf{f}^{\mathrm{T}}(\mathbf{x})\mathbf{A}^{-1}\mathbf{f}(\mathbf{x})$, $\mathbf{A} \in C$, is convex in $\mathbf{A}$ (Groves & Rothenberg 1969). The random variable $\mathbf{M}_{\boldsymbol{\xi}}$ is integrable by Lemma B.7, and $P(\mathbf{M}_{\boldsymbol{\xi}} \in C) = 1$. The conditions of Theorem B.6 hold, hence we see that $\mathrm{E}\{[\mathbf{M}_{\boldsymbol{\xi}}]\} \in C$, therefore $\mathrm{E}\{[\mathbf{M}_{\boldsymbol{\xi}}]\} = nM_\eta$ for some approximate design η. Moreover, $g_{\mathbf{x}}\{\mathrm{E}[\mathbf{M}_{\boldsymbol{\xi}}]\} \leq \mathrm{E}\{g_{\mathbf{x}}[\mathbf{M}_{\boldsymbol{\xi}}]\}$, which implies that

$$\frac{1}{n}\mathbf{f}^{\mathrm{T}}(\mathbf{x})\mathbf{M}_{\eta}^{-1}\mathbf{f}(\mathbf{x}) = \mathbf{f}^{\mathrm{T}}(\mathbf{x})\,\mathrm{E}(\mathbf{M}_{\boldsymbol{\xi}})^{-1}\mathbf{f}(\mathbf{x}) \leq \mathbf{f}^{\mathrm{T}}(\mathbf{x})\,\mathrm{E}[\mathbf{M}_{\boldsymbol{\xi}}^{-1}]\mathbf{f}(\mathbf{x})\,.$$

Note that Lemma B.7 implies that the relevant expectations exist. As $\mathbf{x}$ was arbitrary we can also establish that

$$\frac{1}{n}\max_{\mathbf{x}\in\mathcal{X}} \mathbf{f}^{\mathrm{T}}(\mathbf{x})\mathbf{M}_{\eta}^{-1}\mathbf{f}(\mathbf{x}) \leq \max_{\mathbf{x}\in\mathcal{X}} \mathbf{f}^{\mathrm{T}}(\mathbf{x})\,\mathrm{E}[\mathbf{M}_{\boldsymbol{\xi}}^{-1}]\mathbf{f}(\mathbf{x})\,.$$

However, by definition of η^* the left hand side is at least $\frac{1}{n}\max_{\mathbf{x}\in\mathcal{X}} \mathbf{f}^{\mathrm{T}}(\mathbf{x})\mathbf{M}_{\eta^*}^{-1}\mathbf{f}(\mathbf{x})$. □

Proof of Proposition 4.2 (i) and (ii) are standard. For (iii) note that the information matrix of the deterministic ROAD is equal to $n\mathbf{M}_{\eta^*}$, hence it is a minimax RDS by Lemma B.8.

Proof of Proposition 4.3 For any symmetric positive definite $(p+1)\times(p+1)$ matrix $\mathbf{A}$, define $\phi(\mathbf{A}) = [\max_{\mathbf{x}\in\mathcal{X}} \mathbf{f}^{\mathrm{T}}(\mathbf{x})\mathbf{A}^{-1}\mathbf{f}(\mathbf{x})]^{-1}$. The function ϕ can be shown to be isotonic, i.e. if $\mathbf{A}$ and $\mathbf{B}$ are positive definite matrices with $\mathbf{A}\succeq\mathbf{B}$ in the Loewner partial order then $\phi(\mathbf{A})\geq\phi(\mathbf{B})$. Moreover it is homogeneous, i.e. $\phi(\lambda\mathbf{A})=\lambda\phi(\mathbf{A})$ for all positive definite $\mathbf{A}$ and $\lambda>0$. Note that for a minimax RDS π^* we have by Lemma B.8 that

$$\max_{\mathbf{x}\in\mathcal{X}} \tilde{R}(\mathbf{x},\pi^*) \geq \frac{1}{n\phi[\mathbf{M}_{\eta^*}]}\,. \tag{B.4}$$

Let $\boldsymbol{\xi}_A$ denote the n-run exact design found by Adams rounding of η^*. By Pukelsheim & Rieder (1992) we have $\mathbf{M}_{\boldsymbol{\xi}_A}\succeq(n-K)\mathbf{M}_{\eta^*}$, and also, using isotonicity and homogeneity, that

$$\phi[\mathbf{M}_{\boldsymbol{\xi}_A}] \geq (n-K)\phi[\mathbf{M}_{\eta^*}]\,. \tag{B.5}$$

Using (B.4) and (B.5) we find that max-risk efficiency of $\boldsymbol{\xi}_A$ satisfies

$$\mathrm{eff}(\boldsymbol{\xi}_A;\pi^*) = \frac{\max_{\mathbf{x}\in\mathcal{X}}\tilde{R}(\mathbf{x},\pi^*)}{\max_{\mathbf{x}\in\mathcal{X}}\tilde{R}(\mathbf{x},\boldsymbol{\xi}_A)} \geq \frac{\phi[\mathbf{M}_{\boldsymbol{\xi}_A}]}{n\phi[\mathbf{M}_{\eta^*}]} \geq \frac{(n-K)\phi[\mathbf{M}_{\eta^*}]}{n\phi[\mathbf{M}_{\eta^*}]} \geq 1-\frac{K}{n}\,.$$

B.3 Proofs of results in Section 5

B.3.1 Outline proof of Theorem 5.3

We begin with some definitions. For $\psi\in\mathcal{L}^2(\mathcal{X};\lambda)$ and $\boldsymbol{\xi}=(\mathbf{x}_1,\ldots,\mathbf{x}_n)\in\Xi$, let the *design norm* be $\|\psi\|^2_{\boldsymbol{\xi}}=\sum_{i=1}^n\psi^2(\mathbf{x}_i)=\|\boldsymbol{\psi}_{\boldsymbol{\xi}}\|_2^2$. For an arbitrary random design strategy π, let $I_i\subseteq\mathcal{X}$ denote the support of $\mathbf{x}_i$, and let $N_\pi(\mathbf{x})=\sum_{i=1}^n I(\mathbf{x}\in I_i)$, $\mathbf{x}\in\mathcal{X}$, denote the number of support sets that contain $\mathbf{x}$. Also let $\bar{N}_\pi=\operatorname{ess\,sup}_{\mathbf{x}\in\mathcal{X}}N_\pi(\mathbf{x})$ denote the essential supremum.

Definition B.9. *A random design strategy π is a* uniform random design *strategy if, for $\boldsymbol{\xi} = (\mathbf{x}_1, \ldots, \mathbf{x}_n)$ distributed according to π we have*

$$\mathbf{x}_i \sim \text{Uniform}(I_i)\,, \quad \lambda(I_i) = V\,, \quad \textit{for } i = 1, \ldots, n\,,$$

with $I_i \subseteq \mathcal{X}$ compact sets.

Note that we do not assume independence of the $\mathbf{x}_i$ above. A random translation design strategy $\pi^{\mathrm{RT}}(\bar{\boldsymbol{\xi}}, \mathcal{T})$ is a special case of a uniform random design strategy, with $I_i = \mathbf{c}_i + \mathcal{T}$ and $V = \lambda(\mathcal{T})$. The condition that the I_i are almost disjoint implies that $\bar{N}_\pi \leq 1$. For $\lambda(\mathcal{T}) > 0$ we have $\bar{N}_\pi = 1$.

Lemma B.10. *For a uniform random design strategy with $V > 0$, the design norm of ψ satisfies $\mathrm{E}_{\boldsymbol{\xi}} \|\psi\|_{\boldsymbol{\xi}}^2 \leq (\bar{N}_\pi / V) \int_{\mathcal{X}} \psi^2(\mathbf{x}) d\lambda(\mathbf{x})$.*

Proof of Lemma B.10. Use linearity and write out the expectations explicitly to obtain

$$\begin{aligned}
&\mathrm{E}_{\boldsymbol{\xi}} \left(\sum_{i=1}^n \psi^2(\mathbf{x}_i) \right) = \sum_{i=1}^n V^{-1} \int_{\mathcal{X}} \psi^2(\mathbf{x}) I(\mathbf{x} \in I_i)\, d\lambda(\mathbf{x}) = V^{-1} \int_{\mathcal{X}} \psi^2(\mathbf{x}) N_\pi(\mathbf{x}) d\lambda(\mathbf{x}) \\
&\leq (\bar{N}_\pi / V) \int_{\mathcal{X}} \psi^2(\mathbf{x}) d\lambda(\mathbf{x}) \qquad \text{since } \psi^2(\mathbf{x}) N_\pi(\mathbf{x}) \leq \psi^2(\mathbf{x}) \bar{N}_\pi \text{ almost everywhere.}
\end{aligned}$$

Lemma B.11. *For a uniform random design strategy with $V > 0$, $b(\psi, \pi) = \mathrm{E}[\boldsymbol{\psi}_{\boldsymbol{\xi}}^{\mathrm{T}} \mathbf{K}_{\boldsymbol{\xi}} \boldsymbol{\psi}_{\boldsymbol{\xi}}]$ satisfies*

$$b(\psi, \pi) \leq \frac{\bar{N}_\pi \|\psi\|_2^2}{V} \sup_{\boldsymbol{\xi} \in \mathrm{supp}(\pi)} \lambda_{\max}(\mathbf{K}_{\boldsymbol{\xi}})\,,$$

where $\mathrm{supp}(\pi) \subseteq \Xi$ *denotes the support of the random design strategy π.*

Note that for a random translation design $\pi^{\mathrm{RT}}(\bar{\boldsymbol{\xi}}, \mathcal{T})$ with $\lambda(\mathcal{T}) > 0$ the lemma implies

$$b(\psi, \pi) \leq \frac{\tau^2}{\lambda(\mathcal{T})} \max_{\mathbf{t} \in \mathcal{T}} \lambda_{\max}[\mathbf{K}_{\mathbf{d}(\mathbf{t})}]\,, \quad \psi \in \mathcal{H}\,, \tag{B.6}$$

and the upper bound coincides with the third term in (13). Thus, if it can be shown that the right hand side of (B.6) is the least upper bound over $\psi \in \mathcal{H}$, then Theorem 5.3 is proved. We do so by constructing a family of discrepancy functions $\psi_\epsilon \in \mathcal{H}$, parameterized by ϵ, such that $b(\psi_\epsilon, \pi)$ tends to the bound as $\epsilon \to 0$. A first step towards this is given by the following:

Lemma B.12. *For a non-singular random translation design strategy* $\pi = \pi^{\mathrm{RT}}(\bar{\boldsymbol{\xi}}, \mathcal{T})$ *with* $\lambda(\mathcal{T}) > 0$*, assuming that* $f_0, \ldots, f_p$ *are continuous, there exists a family of functions* $\varphi_\epsilon \in \mathcal{L}^2(\mathcal{X}; \lambda)$*, parameterized by* ϵ*, with* $\|\varphi_\epsilon\|_2^2 = \tau^2$ *and*

$$b(\varphi_\epsilon, \pi) \to \frac{\tau^2}{\lambda(\mathcal{T})} \max_{\mathbf{t} \in \mathcal{T}} \lambda_{\max}[\mathbf{K}_{\mathbf{d}(\mathbf{t})}] \,, \text{ as } \epsilon \to 0 \,,$$

i.e. $b(\varphi_\epsilon, \pi)$ *tends to the upper bound in* (B.6) *as* $\epsilon \to 0$.

However, the functions φ_ϵ constructed in the proof of Lemma B.12 may not be in $\mathcal{H}$, since it is not guaranteed that $\int_{\mathcal{X}} \varphi_\epsilon(\mathbf{x}) \mathbf{f}(\mathbf{x}) d\lambda(\mathbf{x}) = \mathbf{0}_{p+1}$. To obtain functions that are in $\mathcal{H}$, we consider

$$\psi_\epsilon = \varphi_\epsilon - P_{\mathcal{F}}[\varphi_\epsilon] \,.$$

Above, $P_{\mathcal{F}}(\varphi)$ denotes the orthogonal projection of φ into $\mathcal{F} = \mathrm{span}\{f_0, \ldots, f_p\}$, namely $P_{\mathcal{F}}[\varphi] = \sum_{k=0}^{p} \langle \varphi, g_k \rangle g_k$, where $g_0, \ldots, g_p$ denotes an L_2-orthonormal basis of $\mathcal{F}$, e.g. as obtained from $f_0, \ldots, f_p$ by the Gram-Schmidt procedure, and $\langle \cdot, \cdot \rangle$ denotes the $L_2(\mathcal{X}; \lambda)$ inner product.

Lemma B.13. *Again assuming that* $f_0, \ldots, f_p$ *are continuous, the function* ψ_ϵ *is in* $\mathcal{H}$ *and for the non-singular random translation strategy* $\pi = \pi^{\mathrm{RT}}(\bar{\boldsymbol{\xi}}, \mathcal{T})$ *with* $\lambda(\mathcal{T}) > 0$

$$b(\psi_\epsilon, \pi) \to \frac{\tau^2}{\lambda(\mathcal{T})} \max_{\mathbf{t} \in \mathcal{T}} \lambda_{\max}[\mathbf{K}_{\mathbf{d}(\mathbf{t})}] \,, \text{ as } \epsilon \to 0 \,.$$

This is enough to establish that the upper bound is indeed a supremum, and Theorem 5.3 is then proved. The detailed proofs of Lemmas B.11, B.12 and B.13 are given in the next section.

B.3.2 Proofs of Lemmas B.11, B.12 and B.13

Proof of Lemma B.11. For $\boldsymbol{\xi}' \in \mathrm{supp}(\pi)$ and arbitrary $\mathbf{v} \in \mathbb{R}^n$, let $\mathcal{B}(\boldsymbol{\xi}'; \mathbf{v}) = \mathbf{v}^{\mathrm{T}} \mathbf{K}_{\boldsymbol{\xi}'} \mathbf{v}$.

Note

$$b(\psi, \boldsymbol{\xi}') = \mathcal{B}(\boldsymbol{\xi}'; \boldsymbol{\psi}_{\boldsymbol{\xi}'}) , \qquad \mathcal{B}(\boldsymbol{\xi}'; \gamma \mathbf{v}) = \gamma^2 \mathcal{B}(\boldsymbol{\xi}'; \mathbf{v}) . \tag{B.7}$$

For $\boldsymbol{\xi} \sim \pi$, we have that with probability 1 $\boldsymbol{\xi} \in \mathrm{supp}(\pi)$ and (by Lemma B.10) $\|\psi\|_{\boldsymbol{\xi}}^2 = \|\boldsymbol{\psi}_{\boldsymbol{\xi}}\|_{\mathbb{R}^n}^2 < \infty$. Thus we can normalize $\boldsymbol{\psi}_{\boldsymbol{\xi}}$, writing

$$\begin{aligned} \mathcal{B}(\boldsymbol{\xi}; \boldsymbol{\psi}_{\boldsymbol{\xi}}) = \|\psi\|_{\boldsymbol{\xi}}^2 \, \mathcal{B}(\boldsymbol{\xi};\, \boldsymbol{\psi}_{\boldsymbol{\xi}}/\|\psi\|_{\boldsymbol{\xi}}) &\le \|\psi\|_{\boldsymbol{\xi}}^2 \sup_{\{\mathbf{v}: \|\mathbf{v}\|_{\mathbb{R}^n}^2 = 1\}} \mathcal{B}(\boldsymbol{\xi}; \mathbf{v}) \\ &\le \|\psi\|_{\boldsymbol{\xi}}^2 \cdot \lambda_{\max}(\mathbf{K}_{\boldsymbol{\xi}}) \le \|\psi\|_{\boldsymbol{\xi}}^2 \cdot \sup_{\boldsymbol{\xi}' \in \mathrm{supp}(\pi)} \lambda_{\max}(\mathbf{K}_{\boldsymbol{\xi}'}) . \end{aligned}$$

Taking expectations with respect to $\boldsymbol{\xi}$ yields

$$b(\psi, \pi) = \mathrm{E}\{\mathcal{B}(\boldsymbol{\xi}; \boldsymbol{\psi}_{\boldsymbol{\xi}})\} \le \mathrm{E}(\|\psi\|_{\boldsymbol{\xi}}^2) \cdot \sup_{\boldsymbol{\xi}' \in \mathrm{supp}(\pi)} \lambda_{\max}(\mathbf{K}_{\boldsymbol{\xi}'}) \le \frac{\bar{N}_\pi \, \|\psi\|_2^2}{V} \sup_{\boldsymbol{\xi}' \in \mathrm{supp}(\pi)} \lambda_{\max}(\mathbf{K}_{\boldsymbol{\xi}'}) ,$$

where the second inequality follows from Lemma B.10.

Proof of Lemma B.12. First recall that $f_0, \ldots, f_p$ are assumed to be continuous functions and, by the assumption of non-singularity, $\det \mathbf{M}_{\mathbf{d}(\mathbf{t})} \ge \kappa > 0$ for all $\mathbf{t} \in \mathcal{T}$. Hence the map $\mathbf{t} \mapsto \lambda_{\max}(\mathbf{K}_{\mathbf{d}(\mathbf{t})})$ is continuous on $\mathcal{T}$. Since $\mathcal{T}$ is compact, the map attains a (finite) maximum on the compact set $\mathcal{T}$ at some $\mathbf{t}^* \in \arg\max_{\mathbf{t} \in \mathcal{T}} \lambda_{\max}(\mathbf{K}_{\mathbf{d}(\mathbf{t})})$. The vector $\mathbf{t}^*$ is a translation vector defining a maximally bias-sensitive design realization.

Let $\mathcal{T}_\epsilon$ be a a subset of $\mathcal{T}$, parameterized by $\epsilon > 0$, satisfying (i) $\mathcal{T}_\epsilon \to \{\mathbf{t}^*\}$ as $\epsilon \to 0$ and (ii) $\lambda(\mathcal{T}_\epsilon)/\lambda(\mathcal{T}) = \epsilon^q$. E.g., for $0 < \epsilon < 1$ set

$$\mathcal{T}_\epsilon = \{\mathbf{t}^* + \epsilon(\mathbf{t} - \mathbf{t}^*) \,|\, \mathbf{t} \in \mathcal{T}\} .$$

That the above $\mathcal{T}_\epsilon$ is a subset of $\mathcal{T}$ is guaranteed by convexity of $\mathcal{T}$. Note that as $\mathbf{t} \sim \mathrm{Uniform}(\mathcal{T})$, we have $\Pr(\mathbf{t} \in \mathcal{T}_\epsilon) = \lambda(\mathcal{T} \cap \mathcal{T}_\epsilon)/\lambda(\mathcal{T}) = \lambda(\mathcal{T}_\epsilon)/\lambda(\mathcal{T}) = \epsilon^q \lambda(\mathcal{T})/\lambda(\mathcal{T}) = \epsilon^q$.

Define $J_i = \boldsymbol{c}_i + \mathcal{T}_\epsilon$, noting $J_i \subseteq I_i$. As the $I_i = \mathbf{c}_i + \mathcal{T}$ are almost disjoint (by Definition 5.2(ii)), so too are the J_i. Set

$$\varphi_\epsilon(\mathbf{x}) = \sum_{i=1}^{n} \gamma u_i I[\mathbf{x} \in J_i] \tag{B.8}$$

with $\mathbf{u} = (u_1, \ldots, u_n)^{\mathrm{T}}$ a normalized principal eigenvector of $\mathbf{K}_{\mathbf{d}(\mathbf{t}^*)}$. We choose γ to ensure that $\int_{\mathcal{X}} \varphi_\epsilon(\mathbf{x})^2 d\lambda(\mathbf{x}) = \tau^2$. Note $\varphi_\epsilon(\mathbf{x})^2 = \sum_{i=1}^n \gamma^2 u_i^2 I[\mathbf{x} \in J_i] + \sum_{i \neq i'} \gamma^2 u_i u_{i'} I[\mathbf{x} \in J_i \cap J_{i'}]$ and so

$$\int_{\mathcal{X}} \varphi_\epsilon(\mathbf{x})^2 d\lambda(\mathbf{x}) = \gamma^2 \sum_{i=1}^n u_i^2 \lambda(J_i) + \gamma^2 \sum_{i \neq i'} u_i u_{i'} \lambda(J_i \cap J_{i'}) = \gamma^2 \epsilon^q \lambda(\mathcal{T}),$$

using the fact that for $i \neq i'$ J_i and $J_{i'}$ are almost disjoint so $\lambda(J_i \cap J_{i'}) = 0$, and also that $\lambda(J_i) = \lambda(\mathbf{c}_i + \mathcal{T}_\epsilon) = \epsilon^q \lambda(\mathcal{T})$. Thus we set $\gamma^2 = \tau^2 \epsilon^{-q} / \lambda(\mathcal{T})$.

Using (B.8) we have $\varphi_\epsilon(\mathbf{x}_i) = \gamma u_i I[\mathbf{x}_i \in J_i] + \sum_{i' \neq i} \gamma_{i'} u_{i'} I[\mathbf{x}_i \in J_{i'}]$. Since $\mathbf{x}_i \sim \mathrm{Uniform}(I_i)$, for $i \neq i'$ we have that

$$\Pr[\mathbf{x}_i \in J_{i'}] = \lambda[I_i \cap J_{i'}] / \lambda[I_i] \leq \lambda[I_i \cap I_{i'}] / \lambda[I_i] = 0,$$

as I_i and $I_{i'}$ are almost disjoint. Hence $I[\mathbf{x}_i \in J_{i'}] = 0$ almost surely (a.s.), and $\varphi_\epsilon(\mathbf{x}_i) \stackrel{\text{a.s.}}{=} \gamma u_i I[\mathbf{x}_i \in J_i] = \gamma u_i I[\mathbf{t} \in \mathcal{T}_\epsilon]$. Moreover,

$$\boldsymbol{\varphi}_{\mathbf{d}(\mathbf{t})} \stackrel{\text{a.s.}}{=} \gamma I(\mathbf{t} \in \mathcal{T}_\epsilon) \mathbf{u}. \tag{B.9}$$

Hence

$$\begin{aligned} b(\varphi_\epsilon, \pi) &= \mathrm{E}_{\mathbf{t}}\{\mathcal{B}[\mathbf{d}(\mathbf{t}); \boldsymbol{\varphi}_{\mathbf{d}(\mathbf{t})}]\} = \gamma^2 \, \mathrm{E}_{\mathbf{t}}\{\mathcal{B}[\mathbf{d}(\mathbf{t}); \mathbf{u}] I(\mathbf{t} \in \mathcal{T}_\epsilon)\} \quad \text{by (B.7) and (B.9)} \\ &\geq \gamma^2 \left\{ \inf_{\mathbf{t} \in \mathcal{T}_\epsilon} \mathcal{B}[\mathbf{d}(\mathbf{t}); \mathbf{u}] \right\} P(\mathbf{t} \in \mathcal{T}_\epsilon) \\ &= \frac{\tau^2}{\lambda(\mathcal{T})} \inf_{\mathbf{t} \in \mathcal{T}_\epsilon} \mathcal{B}[\mathbf{d}(\mathbf{t}); \mathbf{u}] \to \frac{\tau^2}{\lambda(\mathcal{T})} \mathcal{B}[\mathbf{d}(\mathbf{t}^*); \mathbf{u}] = \frac{\tau^2}{\lambda(\mathcal{T})} \lambda_{\max}[\mathbf{K}_{\mathbf{d}(\mathbf{t}^*)}], \quad \text{as } \epsilon \to 0, \end{aligned}$$

where in the last line we have used continuity of $\mathcal{B}[\mathbf{d}(\mathbf{t}), \mathbf{u}]$ in $\mathbf{t}$, which follows from continuity of $\mathbf{K}_{\mathbf{d}(\mathbf{t})}$, and the fact that $\max_{\mathbf{t} \in \mathcal{T}_\epsilon} \|\mathbf{t} - \mathbf{t}^*\| = \epsilon \max_{\mathbf{t} \in \mathcal{T}} \|\mathbf{t} - \mathbf{t}^*\| \to 0$ since $\mathcal{T}$ is compact and so bounded. Hence,

$$\lim_{\epsilon \to 0} b(\varphi_\epsilon, \pi) \geq \frac{\tau^2}{\lambda(\mathcal{T})} \lambda_{\max}[\mathbf{K}_{\mathbf{d}(\mathbf{t}^*)}],$$

thus completing the proof of Lemma B.12.

Before proving Lemma B.13, we state some definitions and results that are standard in the mathematical analysis literature (e.g. Rynne & Youngson 2008, p.129). Let V be a vector space over a field $\mathbb{F}$. Then $\|\cdot\| : V \to [0, \infty)$ is a *seminorm* if, for all $\lambda \in \mathbb{F}$, $v, w \in V$,

1. $\|\lambda v\| = |\lambda| \, \|v\|$ (scalability)

2. $\|v + w\| \le \|v\| + \|w\|$ (triangle inequality)

3. $\|v\| \ge 0$ (non-negativity).

Recall that a bilinear form $A : V \times V \to \mathbb{R}$ is positive semi-definite if, for all $v \in V$, $A(v, v) \ge 0$, and symmetric if $A(u, v) = A(v, u)$ for all $u, v \in V$.

Lemma B.14. *If A is a positive semi-definite symmetric bilinear form, then $\|\cdot\|$ defined by $\|v\| = \sqrt{A(v, v)}$ is a seminorm.*

The triangle inequality is established using the Cauchy-Schwarz inequality, whose usual proof is still valid, as a stepping stone.

Proof of Lemma B.13. First note that by the definition, ψ_ϵ is L_2-orthogonal to $\mathbf{f}$ and to $P_{\mathcal{F}}[\varphi_\epsilon]$. Moreover, by Pythagoras's theorem,

$$\|\psi_\epsilon\|_2^2 = \|\varphi_\epsilon\|_2^2 - \|P_{\mathcal{F}}[\varphi_\epsilon]\|_2^2 \le \|\varphi_\epsilon\|_2^2 = \tau^2 \, ,$$

and so ψ_ϵ is indeed in $\mathcal{H}$.

Note that

$$\begin{aligned} \epsilon^{-q/2} \langle \varphi_\epsilon, g_k \rangle &= \epsilon^{-q/2} \gamma \int_{\mathcal{X}} \left\{ \sum_{i=1}^n u_i I(\mathbf{x} \in J_i) \right\} g_k(\mathbf{x}) d\lambda(\mathbf{x}) \\ &= \epsilon^{-q/2} \gamma \sum_{i=1}^n u_i \int_{J_i} g_k(\mathbf{x}) d\lambda(\mathbf{x}) \\ &= \frac{\tau}{\epsilon^q \lambda(\mathcal{T})^{1/2}} \sum_{i=1}^n u_i \int_{\mathbf{c}_i + \mathcal{T}_\epsilon} g_k(\mathbf{x}) d\lambda(\mathbf{x}) \to \tau \lambda(\mathcal{T})^{1/2} \sum_{i=1}^n u_i g_k(\mathbf{c}_i + \mathbf{t}^*) \, , \end{aligned}$$

using the assumption that $f_0, \ldots, f_p$ and so also $g_0, \ldots, g_p$ are continuous. Consequently, $\langle \varphi_\epsilon, g_k \rangle = O(\epsilon^{q/2})$ as $\epsilon \to 0$.

Note that by the fact above that $\langle \varphi_\epsilon, g_k \rangle = O(\epsilon^{q/2})$, together with the triangle inequality,

$$\|\psi_\epsilon - \varphi_\epsilon\|_2 = \|P_{\mathcal{F}}[\varphi_\epsilon]\|_2 \le \sum_{k=0}^p |\langle \varphi_\epsilon, g_k \rangle| \, \|g_k\| = O(\epsilon^{q/2}) \, , \tag{B.10}$$

and so, for small ϵ, φ_ϵ is well approximated by ψ_ϵ.

Let $\|\cdot\|_{EB}$ be defined, for Ψ in $\mathcal{L}^2(\mathcal{X};\lambda)$, by $\|\Psi\|_{EB}^2 = \mathrm{E}_{\boldsymbol{\xi}}\,\mathcal{B}(\boldsymbol{\xi};\Psi) = b(\Psi,\pi)$. Define $A : \mathcal{L}^2(\mathcal{X};\lambda)\times\mathcal{L}^2(\mathcal{X};\lambda) \to [0,\infty)$ by $A(\varphi^{(1)},\varphi^{(2)}) = \mathrm{E}_{\boldsymbol{\xi}}\{\boldsymbol{\varphi}_{\boldsymbol{\xi}}^{(1)\mathrm{T}}\mathbf{K}_{\boldsymbol{\xi}}\boldsymbol{\varphi}_{\boldsymbol{\xi}}^{(2)}\}$, noting that $\|\psi\|_{EB} = A(\psi,\psi)^{1/2} = b(\psi,\pi)^{1/2}$. We claim that A is a positive semi-definite symmetric bilinear form, and so $\|\cdot\|_{EB}$ is a seminorm by Lemma B.14. We have already shown with Lemma B.11 that $\|\cdot\|_{EB}$ is dominated by $\|\cdot\|_2$, since

$$\|\cdot\|_{EB} \le \left\{\frac{1}{\lambda(\mathcal{T})^{1/2}}\sup_{\mathbf{t}\in\mathcal{T}}\lambda_{\max}^{1/2}\left(\mathbf{K}_{\mathbf{d}(\mathbf{t})}\right)\right\}\|\cdot\|_2 \ .$$

Combining this with (B.10), we have that $\|P_{\mathcal{F}}[\varphi_\epsilon]\|_{EB} \to 0$ as $\epsilon \to 0$. By the triangle inequality for $\|\cdot\|_{EB}$,

$$\|\mathcal{H}\|_{EB} \ge \|\psi_\epsilon\|_{EB} \ge \|\varphi_\epsilon\|_{EB} - \|P_{\mathcal{F}}[\varphi_\epsilon]\|_{EB} \to \|\mathcal{H}\|_{EB} \ ,$$

where $\|\mathcal{H}\|_{EB} = \tau\lambda(\mathcal{T})^{-1/2}\sup_{\mathbf{t}\in\mathcal{T}}\lambda_{\max}^{1/2}(\mathbf{K}_{\mathbf{d}(\mathbf{t})})$ (the square root of the supremum we are trying to approach). Therefore in fact, $\psi_\epsilon \in \mathcal{H}$, with

$$\lim_{\epsilon\to 0}\|\psi_\epsilon\|_{EB} = \|\mathcal{H}\|_{EB} \ ,$$

and the result is proved.

B.3.3 Constraints on hypercuboidal random translation designs

Proof of Lemma 5.4. Condition (i) is equivalent to the statement that $\mathbf{c}_i + [-\delta/2,\delta/2]^q \subseteq \mathcal{X}$, $i = 1,\dots,n$, since $\mathbf{c}_i + \mathcal{T} \subseteq [-1,1]^q$ if and only if $c_{ij} + \delta/2 \le 1$ and $c_{ij} - \delta/2 \ge -1$ for all $i = 1,\dots,n$, $j = 1,\dots,q$; equivalently, if and only if $-1+\delta/2 \le c_{ij} \le 1-\delta/2$.

Condition (ii) is necessary and sufficient for I_i and $I_{i'}$, $i \ne i'$, to be almost disjoint, where $I_i = \mathbf{c}_i + [-\delta/2,\delta/2]^q$. To see this, let $\mathbf{a}$ and $\mathbf{b}$ be vectors in $\mathbb{R}^q$, and $I = \mathbf{a} + [-\delta/2,\delta/2]^q$, $I' = \mathbf{b} + [-\delta/2,\delta/2]^q$. We show that: (a) $I \cap I'$ is empty if $\|\mathbf{b}-\mathbf{a}\|_\infty > \delta$, (b) $\lambda(I\cap I') = 0$ if $\|\mathbf{b}-\mathbf{a}\|_\infty = \delta$, and (c) $\lambda(I\cap I') > 0$ if $\|\mathbf{b}-\mathbf{a}\|_\infty < \delta$. Hence I and I' are almost disjoint if and only if $\|\mathbf{b}-\mathbf{a}\|_\infty \geq \delta$. Hence $I_i = \mathbf{c}_i + [-\delta/2, \delta/2]^q$, $i = 1, \ldots, n$, are pairwise almost disjoint if and only if $\|\mathbf{c}_i - \mathbf{c}_{i'}\| \geq \delta$ for all $i \neq i$, or equivalently if and only if $\min_{i,i'} \|\mathbf{c}_i - \mathbf{c}_{i'}\|_\infty \geq \delta$. This is enough to prove the result.

To see (a), (b) and (c) above note that $\mathbf{x} \in I \cap I'$ if and only if

$$a_j - \delta/2 \leq x_j \leq a_j + \delta/2, \quad b_j - \delta/2 \leq x_j \leq b_j + \delta/2 \qquad \text{for } j = 1, \ldots, q\,,$$

which is equivalent to the condition that $\max\{a_j, b_j\} - \delta/2 \leq x_j \leq \min\{a_j, b_j\} + \delta/2$ for all j. In particular, $I \cap I'$ is non-empty if and only if $\min\{a_j, b_j\} + \delta/2 \geq \max\{a_j, b_j\} - \delta/2$ for all j, or equivalently $|b_j - a_j| \leq \delta$ for all j, i.e. $\|\mathbf{b}-\mathbf{a}\|_\infty \leq \delta$. In this case

$$I \cap I' = \prod_{j=1}^{q} \left[\max\{a_j, b_j\} - \frac{\delta}{2}, \min\{a_j, b_j\} + \frac{\delta}{2}\right]$$

and $\lambda(I \cap I') = \prod_{j=1}^{q} [\delta - |b_j - a_j|]$. Thus if $\|\mathbf{b}-\mathbf{a}\|_\infty = \max_j |b_j - a_j| = \delta$, then $|b_j - a_j| = \delta$ for some j and $\lambda(I \cap I') = 0$. If $\|\mathbf{b}-\mathbf{a}\|_\infty = d < \delta$ then $\lambda(I \cap I') \geq [\delta - d]^q > 0$.

B.3.4 Proof of Proposition 5.1

We show that there is a family of discrepancy functions $\psi_\delta \in \mathcal{H}$, parameterized by δ, such that for all $u \geq 0$ we have $S(\boldsymbol{\theta}_\delta, \boldsymbol{\xi}, u) \to 1$ as $\delta \to 0$, where $\boldsymbol{\theta}_\delta = (\psi_\delta, \boldsymbol{\beta}, \sigma^2)$. This is enough to establish the result.

Let $\phi_\delta(\mathbf{x}) = \sum_{i=1}^{n} \frac{\tau}{\delta^{q/2}} b_i I[\|\mathbf{x} - \mathbf{x}_i\|_\infty \leq \delta/2]$, $b_i \in \mathbb{R}$, $i = 1, \ldots, n$, $\sum_{i=1}^{n} b_i^2 = 1$, and let

$$\psi_\delta(\mathbf{x}) = \phi_\delta(\mathbf{x}) - \sum_{j=0}^{p} \frac{\langle g_j, \phi \rangle_2}{\|g_j\|_2^2} g_j(\mathbf{x}) \tag{B.11}$$

denote the orthogonal component of the L_2-projection of ϕ_δ into $\text{span}(f_0, \ldots, f_p)$. Above, $g_0, \ldots, g_p$ denotes an orthogonal basis of $\text{span}(f_0, \ldots, f_p)$. This construction ensures that ψ_δ is orthogonal to each f_j, a necessary condition for ψ_δ to be in $\mathcal{H}$. By arguments in the proof of Lemma 5.4, provided that $\delta < \min_{i,j} \|\mathbf{x}_i - \mathbf{x}_j\|_\infty$, the sets $S_i = \{\mathbf{x} : \|\mathbf{x} - \mathbf{x}_i\|_\infty < \delta/2\}$ do not overlap and so $\|\phi_\delta\|_2^2 = \sum_{i=1}^{n} \int_{S_i} (\tau^2 b_i^2 / \delta^q) d\lambda(\mathbf{x}) = \tau^2$. By Pythagoras' theorem, we have that $\tau^2 = \|\phi_\delta\|_2^2 = \|\psi_\delta\|_2^2 + \|\phi_\delta - \psi_\delta\|_2^2$ and so $\|\psi_\delta\|_2^2 \le \tau^2$. Hence ψ_δ is in $\mathcal{H}$. By the same argument as in the proof of Lemma B.13, assuming that the f_j are continuous we have $\langle g_j, \phi_\delta \rangle = O(\delta^{q/2})$ as $\delta \to 0$. Hence, if we let $\boldsymbol{\psi}_{\boldsymbol{\xi}}^\delta = (\psi_\delta(\mathbf{x}_1), \ldots, \psi_\delta(\mathbf{x}_n))^{\mathrm{T}}$ be the vector containing evaluations of the model discrepancy function at the design points, then

$$\boldsymbol{\psi}_{\boldsymbol{\xi}}^\delta = (\phi_\delta(\mathbf{x}_1), \ldots, \phi_\delta(\mathbf{x}_n))^{\mathrm{T}} + O(\delta^{q/2}) = \frac{\tau}{\delta^{q/2}}\mathbf{b} + O(\delta^{q/2}) \quad \text{as } \delta \to 0\,,$$

where $\mathbf{b} = (b_1, \ldots, b_n)^{\mathrm{T}}$. Moreover, $\|\psi_\delta\|_2^2 \to \|\phi_\delta\|_2^2 = \tau^2$.

Now note that for $\boldsymbol{\theta}_\delta$ we have

$$\begin{aligned}
\hat{\boldsymbol{\beta}} &= \mathbf{M}_{\boldsymbol{\xi}}^{-1}\mathbf{F}_{\boldsymbol{\xi}}^{\mathrm{T}}\mathbf{y} = \boldsymbol{\beta} + \mathbf{M}_{\boldsymbol{\xi}}^{-1}\mathbf{F}_{\boldsymbol{\xi}}^{\mathrm{T}}\boldsymbol{\psi}_{\boldsymbol{\xi}}^\delta + \mathbf{M}_{\boldsymbol{\xi}}^{-1}\mathbf{F}_{\boldsymbol{\xi}}^{\mathrm{T}}\boldsymbol{\epsilon} \\
&= \boldsymbol{\beta} + \frac{\tau}{\delta^{q/2}}\mathbf{M}_{\boldsymbol{\xi}}^{-1}\mathbf{F}_{\boldsymbol{\xi}}^{\mathrm{T}}\mathbf{b} + O_p(1)
\end{aligned}$$

and

$$\left\|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right\|_2^2 = \frac{\tau^2}{\delta^q}\mathbf{b}^{\mathrm{T}}\mathbf{F}_{\boldsymbol{\xi}}\mathbf{M}_{\boldsymbol{\xi}}^{-2}\mathbf{F}_{\boldsymbol{\xi}}^{\mathrm{T}}\mathbf{b} + O_p(\delta^{-q/2})\,.$$

Note also that

$$\tau^2\mathbf{b}^{\mathrm{T}}\mathbf{F}_{\boldsymbol{\xi}}\mathbf{M}_{\boldsymbol{\xi}}^{-2}\mathbf{F}_{\boldsymbol{\xi}}^{\mathrm{T}}\mathbf{b} \ge \tau^2\lambda_{\min}(\mathbf{F}_{\boldsymbol{\xi}}\mathbf{M}_{\boldsymbol{\xi}}^{-2}\mathbf{F}_{\boldsymbol{\xi}}^{\mathrm{T}}) = \tau^2/\lambda_{\max}(\mathbf{M}_{\boldsymbol{\xi}}) > 0\,, \tag{B.12}$$

because λ is an eigenvalue of $\mathbf{F}_{\boldsymbol{\xi}}\mathbf{M}_{\boldsymbol{\xi}}^{-2}\mathbf{F}_{\boldsymbol{\xi}}^{\mathrm{T}}$ if and only if $1/\lambda$ is an eigenvalue of $\mathbf{M}_{\boldsymbol{\xi}}$ (which we assume is non-singular). The loss is

$$\begin{aligned}
\ell(\boldsymbol{\theta}, \hat{\boldsymbol{\alpha}}) &= \int_{\mathcal{X}} [\mu(\mathbf{x}) - \mathbf{f}^{\mathrm{T}}(\mathbf{x})\hat{\boldsymbol{\beta}}]^2 d\lambda(\mathbf{x}) = (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})^{\mathrm{T}}\mathbf{A}(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}) + \|\psi\|_2^2\,, \\
&\ge \left\|\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}\right\|_2^2 \lambda_{\min}(\mathbf{A}) + \|\psi\|_2^2\,,
\end{aligned}$$

where $\mathbf{A} = \int_{\mathcal{X}} \mathbf{f}(\mathbf{x})\mathbf{f}^{\mathrm{T}}(\mathbf{x})d\lambda(\mathbf{x})$. We assume as a mild regularity condition on the model that

$\lambda_{\min}(\mathbf{A}) > 0$. Hence

$$\begin{aligned} S(\boldsymbol{\theta}_\delta, \boldsymbol{\xi}, u) &= \Pr[\ell(\boldsymbol{\theta}_\delta, \hat{\boldsymbol{\alpha}}) > u] \\ &\geq \Pr\left[\left\|\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}\right\|_2^2 > \frac{u - \|\psi_\delta\|_2^2}{\lambda_{\min}(\mathbf{A})}\right] \\ &\geq \Pr\left[\delta^q \left\|\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}\right\|_2^2 > \frac{\delta^q(u - \|\psi_\delta\|_2^2)}{\lambda_{\min}(\mathbf{A})}\right] \\ &\geq \Pr\left[\tau^2 \mathbf{b}^{\mathrm{T}} \mathbf{F}_{\boldsymbol{\xi}} \mathbf{M}_{\boldsymbol{\xi}}^{-2} \mathbf{F}_{\boldsymbol{\xi}}^{\mathrm{T}} \mathbf{b} > O_p(\delta^{q/2})\right] \to 1 \text{ as } \delta \to 0\,, \end{aligned}$$

because $\tau^2 \mathbf{b}^{\mathrm{T}} \mathbf{F}_{\boldsymbol{\xi}} \mathbf{M}_{\boldsymbol{\xi}}^{-2} \mathbf{F}_{\boldsymbol{\xi}}^{\mathrm{T}} \mathbf{b}$ is a fixed positive quantity, by (B.12). This completes the proof.

B.4 Further details for Section 5.3

B.4.1 Refinement of the discretization $\tilde{\mathcal{T}}$

As mentioned in Section 5.3.1, we allow iterative refinement of the discretization set using an approach similar to Pronzato & Pázman (2013, p.311). This leads to the following overall algorithm for numerical optimization of the strategy $\pi^{\mathrm{H}}(\bar{\boldsymbol{\xi}}, \delta)$. Write $\hat{\Psi}(\bar{\boldsymbol{\xi}}, \delta; \tilde{\mathcal{T}}) = \hat{\Psi}(\bar{\boldsymbol{\xi}}, \delta)$ and $\hat{\Psi}_2(\tilde{\boldsymbol{\xi}}, \tilde{\delta}; \tilde{\mathcal{T}}) = \hat{\Psi}_2(\tilde{\boldsymbol{\xi}}, \tilde{\delta})$ to emphasize the dependence of the approximate objective functions on the discretization $\tilde{\mathcal{T}}$.

Require: initial discretization $\tilde{\mathcal{T}}^{(1)} = \{\tilde{\mathbf{t}}_1, \dots \tilde{\mathbf{t}}_M\} \subseteq [-1,1]^q$, tolerance $\epsilon > 0$

Set $l = 1$

repeat

Compute $(\tilde{\boldsymbol{\xi}}^{(l)}, \tilde{\delta}^{(l)}) \in \arg\min_{\tilde{\boldsymbol{\xi}},\tilde{\delta}} \hat{\Psi}_2(\tilde{\boldsymbol{\xi}}, \tilde{\delta}; \tilde{\mathcal{T}}^{(l)})$ via multiple random initializations of a co-ordinate descent algorithm

Compute $\tilde{\mathbf{t}}_{M+l} \in \arg\max_{\tilde{\mathbf{t}}\in[-1,1]^q} \lambda_{\max}(\mathbf{K}_{\mathbf{d}(\frac{\delta}{2}\tilde{\mathbf{t}})})$ via co-ordinate ascent

if $\frac{\tau^2}{\delta^q} \max_{\tilde{\mathbf{t}}\in\mathcal{T}^{(l)}} \lambda_{\max}(\mathbf{K}_{\mathbf{d}(\frac{\delta}{2}\tilde{\mathbf{t}})}) > \frac{\tau^2}{\delta^q}\lambda_{\max}(\mathbf{K}_{\mathbf{d}(\frac{\delta}{2}\tilde{\mathbf{t}}_{M+l})}) - \epsilon$ **then**

return $(\tilde{\boldsymbol{\xi}}^l, \tilde{\delta}^{(l)})$ as an ϵ-optimal strategy

else

set $\tilde{\mathcal{T}}^{(l+1)} = \tilde{\mathcal{T}}^{(l)} \cup \{\tilde{\mathbf{t}}_{M+l}\}$

update $l \leftarrow l + 1$

end if

until convergence

In practice, for our examples if $\mathcal{T}^{(1)}$ contains all factorial points of the form $(\pm 1, \ldots, \pm 1) \in [-1, 1]^q$ then further refinement of the discretization is rarely needed.

B.4.2 Heuristic strategy in Section 5.3.3

The heuristic mean design $\bar{\boldsymbol{\xi}}_\delta$ used for the $q = 3$ factor, $n = 12$ run problem is given in Table 2. Setting $\delta = 0$ gives an approximately V-optimal exact design.

B.5 Results on the attained conditional risk

We now demonstrate that, as claimed in Section 6, for the minimax random design strategy π_{mM} in Section 4.1.4 we have: (i) a tight lower bound on the probability is $\Pr[R(\boldsymbol{\theta}, \boldsymbol{\xi}) < \max_{\boldsymbol{\theta}' \in \Theta} R(\boldsymbol{\theta}', \boldsymbol{\xi}_{\mathrm{mM}})] \geq 0.684$; and (ii) the probability that $R(\boldsymbol{\theta}, \boldsymbol{\xi}) > \max_{\boldsymbol{\theta} \in \Theta} R(\boldsymbol{\theta}, \boldsymbol{\xi}_{\mathrm{mM}})$ is at most 0.118.

Recall that for $\boldsymbol{\theta} = (\mathbf{x}, \sigma^2, \boldsymbol{\beta}) \in \mathcal{X} \times [\underline{\sigma}^2, \bar{\sigma}^2] \times B$ the conditional risk function is $R(\boldsymbol{\theta}, \boldsymbol{\xi}) = \sigma^2 \tilde{R}(\mathbf{x}, \boldsymbol{\xi})$ and $\max_{\boldsymbol{\theta}' \in \Theta} R(\boldsymbol{\theta}', \boldsymbol{\xi}) = \bar{\sigma}^2 \max_{\mathbf{x}' \in \mathcal{X}} \tilde{R}(\mathbf{x}', \boldsymbol{\xi})$. Table 3 enumerates the value of $\tilde{R}(\mathbf{x}', \boldsymbol{\xi})$ for $\mathbf{x}' \in \mathcal{X}$, $\boldsymbol{\xi} \in \operatorname{supp}(\pi_{\mathrm{mM}})$ to 3 d.p.. Also note that $\max_{\mathbf{x}' \in \mathcal{X}} \tilde{R}(\mathbf{x}', \boldsymbol{\xi}_{\mathrm{mM}}) = 2.75$. Hence, using Table 3, we obtain the following tight bounds as claimed:

$$
\begin{aligned}
\text{(i)} \quad & \Pr\left[R(\boldsymbol{\theta}, \boldsymbol{\xi}) < \max_{\boldsymbol{\theta}' \in \Theta} R(\boldsymbol{\theta}', \boldsymbol{\xi}_{\mathrm{mM}})\right] = \Pr\left[\tilde{R}(\mathbf{x}, \boldsymbol{\xi}) < \frac{2.75\bar{\sigma}^2}{\sigma^2}\right] \\
& \geq \min_{\mathbf{x}' \in \mathcal{X}} \Pr[\tilde{R}(\mathbf{x}', \boldsymbol{\xi}) < 2.75] = 0.684; \\
\text{(ii)} \quad & \Pr\left[R(\boldsymbol{\theta}, \boldsymbol{\xi}) > \max_{\boldsymbol{\theta}' \in \Theta} R(\boldsymbol{\theta}', \boldsymbol{\xi}_{\mathrm{mM}})\right] = \Pr\left[\tilde{R}(\mathbf{x}, \boldsymbol{\xi}) > \frac{2.75\bar{\sigma}^2}{\sigma^2}\right] \\
& \leq \max_{\mathbf{x}' \in \mathcal{X}} \Pr[\tilde{R}(\mathbf{x}', \boldsymbol{\xi}) > 2.75] = 0.118.
\end{aligned}
$$

$\boldsymbol{\xi}_L$			$\boldsymbol{\xi}_{\text{bad}}$		
x_1	x_2	x_3	x_1	x_2	x_3
−1	−1	−1	−1	−1	0
−1	−1	1	−1	−1	0
−1	0	0	−1	−1	0
−1	0	0	−1	−1	0
−1	1	−1	−1	−1	0
−1	1	1	−1	0	1
0	−1	−1	0	−1	−1
0	−1	0	0	0	−1
0	0	−1	0	0	−1
0	0	1	0	0	−1
0	1	−1	1	−1	1
0	1	0	1	−1	1
1	−1	−1	1	0	0
1	−1	0	1	0	0
1	−1	1	1	0	1
1	0	−1	1	1	−1
1	0	1	1	1	−1
1	1	−1	1	1	0
1	1	0	1	1	0
1	1	1	1	1	1

Table 1: An L-optimal design, $\boldsymbol{\xi}_L^*$, and a poor design, $\boldsymbol{\xi}_{\text{bad}}$, for the example in Section 3.2.

$\bar{\boldsymbol{\xi}}_\delta$		
x_1	x_2	x_3
$-1+\frac{\delta}{2}$	$-1+\frac{\delta}{2}$	$-1+\frac{\delta}{2}$
$-1+\frac{\delta}{2}$	$-1+\frac{\delta}{2}$	$1-\frac{\delta}{2}$
$-1+\frac{\delta}{2}$	0.313	-0.313
$-1+\frac{\delta}{2}$	$1-\frac{\delta}{2}$	$1-\frac{\delta}{2}$
-0.166	$1-\frac{\delta}{2}$	$-1+\frac{\delta}{2}$
$-0.093-\frac{\delta}{2}$	-0.092	0.093
$-0.093+\frac{\delta}{2}$	-0.092	0.093
0.310	$-1+\frac{\delta}{2}$	-0.311
0.313	0.312	$1-\frac{\delta}{2}$
$1-\frac{\delta}{2}$	$-1+\frac{\delta}{2}$	$1-\frac{\delta}{2}$
$1-\frac{\delta}{2}$	-0.170	$-1+\frac{\delta}{2}$
$1-\frac{\delta}{2}$	$1-\frac{\delta}{2}$	0.168

Table 2: Heuristic $\bar{\boldsymbol{\xi}}_\delta$ used in Section 5.3.3

	$\mathbf{x}'^{\mathrm{T}}$	$(-1,-1)$	$(0,-1)$	$(1,-1)$	$(-1,0)$	$(0,0)$	$(1,0)$	$(-1,1)$	$(0,1)$	$(1,1)$
$\boldsymbol{\xi}$	$\pi(\boldsymbol{\xi})$	$\tilde{R}(\mathbf{x},\boldsymbol{\xi})$								
$\boldsymbol{\xi}_1$	0.211	1.000	1.000	1.000	1.000	2.750	2.000	1.000	2.000	1.000
$\boldsymbol{\xi}_2$	0.118	1.000	1.000	5.667	1.667	1.000	1.000	1.000	1.667	1.000
$\boldsymbol{\xi}_3$	0.118	1.000	1.667	1.000	1.000	1.000	1.667	5.667	1.000	1.000
$\boldsymbol{\xi}_4$	0.118	1.000	1.667	1.000	1.667	1.000	1.000	1.000	1.000	5.667
$\boldsymbol{\xi}_5$	0.105	1.000	2.750	1.000	2.000	1.000	1.000	1.000	2.750	1.000
$\boldsymbol{\xi}_6$	0.105	1.000	2.000	1.000	2.750	1.000	2.750	1.000	1.000	1.000
$\boldsymbol{\xi}_7$	0.105	1.000	2.000	1.000	2.000	2.750	1.000	1.000	1.000	1.000
$\boldsymbol{\xi}_8$	0.118	5.667	1.000	1.000	1.000	1.000	1.667	1.000	1.667	1.000
$\Pr[\tilde{R}(\mathbf{x}',\boldsymbol{\xi}) < 2.75]$		0.882	0.895	0.882	0.895	0.684	0.895	0.882	0.895	0.882
$\Pr[\tilde{R}(\mathbf{x}',\boldsymbol{\xi}) > 2.75]$		0.118	0.000	0.118	0.000	0.000	0.000	0.118	0.000	0.118

Table 3: Detailed computation of the probability that the attained conditional risk is smaller than (or greater than) $\max_{\boldsymbol{\theta}'\in\Theta} R(\boldsymbol{\theta}',\boldsymbol{\xi}_{\mathrm{mM}})$, for the minimax RDS in Section 4.1.4.

References

Berger, J. O. (1985), *Statistical decision theory and Bayesian analysis*, 2nd edn, Springer, New York.

Berkelaar, M., Eikland, K. & Notebaert, P. (2004), 'lpsolve: Open source (mixed-integer) linear programming system, version 5.1.0.0'.
URL: *http://lpsolve.sourceforge.net*

Ferguson, T. S. (1967), *Mathematical statistics: A decision theoretic approach*, Academic press, London.

Groves, T. & Rothenberg, T. (1969), 'A note on the expected value of an inverse matrix', *Biometrika* **56**, 690–691.

Konis, K. (2014), 'lpsolveapi: R interface for lpsolve version 5.5.2.0', R package.
URL: *https://cran.r-project.org/web/packages/lpSolveAPI/index.html*

Pronzato, L. & Pázman, A. (2013), 'Design of experiments in nonlinear models', *Lecture notes in Statistics* **212**.

Pukelsheim, F. & Rieder, S. (1992), 'Efficient rounding of approximate designs', *Biometrika* **79**, 763–770.

Rynne, B. P. & Youngson, M. A. (2008), *Linear functional analysis*, 2nd edn, Springer, London.

Thie, P. R. & Keough, G. E. (2011), *An Introduction to Linear Programming and Game Theory*, Wiley, Hoboken.

Wu, C.-F. (1981), 'On the robustness and efficiency of some randomized designs', *Ann. Statist.* **9**, 1168–1177.