Journal of Statistical Planning and Inference 205 (2020) 1–9

Contents lists available at ScienceDirect

Journal of Statistical Planning and Inference

journal homepage: www.elsevier.com/locate/jspi

Small screening design when the overall variance is unknown

Jiaya Peng *, Dennis K.J. Lin

Department of Statistics, The Pennsylvania State University, State College, PA 16802, United States

ARTICLE INFO

Article history:
Received 9 December 2017
Received in revised form 13 March 2019
Accepted 1 April 2019
Available online 29 May 2019

Keywords:
Center points
Estimation of variance
Kurtosis
Robustness
Test of significance

ABSTRACT

Consider the problem of a design for estimating all main effects and the overall error variance. The conventional choice is a saturated orthogonal array supplemented by several center points (as popularized in Response Surface Methodology). We propose an alternative design — the projection of a larger Hadamard matrix. In this article, we prove that the proposed design is optimal for estimating error variance, and thus is preferable over the conventional choice. Under various common error distributions, theoretical values of $\mathrm{Var}(\hat{\sigma}^2)$ are evaluated to illustrate our theory. Simulation results are provided to demonstrate that the proposed design achieves a more reliable estimate of error variance, as well as a reliable significant test. Furthermore, we study the optimal follow-up plan. Relevant optimality theories are established and a convenient construction method for optimal follow-up design is proposed.

1. Introduction

To provide extra degrees of freedom for estimating the error variance (σ^2), one traditional design strategy is to add center points. Under a first-order main-effect model with k experimental factors, suppose one is asked to conduct an n-run design where n is slightly greater than $k+1$. The conventional wisdom is to perform a $(k+1)\times k$ orthogonal main-effect design, e.g., a Plackett–Burman (Plackett and Burman, 1946) design, and then add $(n-k-1)$ center points. In this article, we show that adding center points may not be a favorable plan for estimating σ^2. We show that to perform the $n\times k$ design as a whole, a better and in fact optimal plan is to choose k column from an n-run orthogonal main-effect design. Optimality theorems are derived and illustrative examples are given.

As a motivative example, consider a study involving seven factors $x_1,\ldots,x_7$, via a linear first-order model $y_i = \beta_0+\beta_1x_{i1}+\beta_2x_{i2}+\cdots+\beta_7x_{i7}+\epsilon_i$, where ϵ_i's are independent and identically distributed (i.i.d.) with mean 0 and variance σ^2. Here we have eight effect parameters $\beta_0,\beta_1,\ldots,\beta_7$, as well as the error variance σ^2 to estimate. For estimation of $\boldsymbol{\beta}=(\beta_0,\beta_1,\ldots,\beta_7)^\top$, an eight-run experiment will be sufficient. Suppose we decide to add four additional runs for estimating σ^2, thus a total of 12 runs is called for.

How to design such an experiment in 12 runs? The conventional wisdom is to add four center points, to a 2^{7-4} fractional factorial design, or here we call it an 8×7 "H-design"; see Design A in Table 1(a). We propose an alternative plan: choose any 7 columns from a 12×11 H-design; see Design B in Table 1(b). Throughout, an m-run H-design refers to any $m\times(m-1)$ matrix $\mathbf{H}$ such that $[\mathbf{1},\mathbf{H}]$ forms an Hadamard matrix. (Here m is any multiplier of 4; $\mathbf{1}$ is a column of $+1$'s; the two levels in the Hadamard matrix are labeled as "+ " and "−".) As will be shown, Design B is more preferable than Design A in terms of estimating σ^2. Explicitly, $\mathrm{Var}(\hat{\sigma}^2)$ is smaller under Design B, where $\hat{\sigma}^2$ is the estimate of σ^2.

The rest of the paper is organized as follows. In Section 2, we present the explicit problem formulation, and establish the optimality of Design B in estimating σ^2. Under various common distributions, theoretical values of $\mathrm{Var}(\hat{\sigma}^2)$ have been

* Corresponding author.
E-mail addresses: pengjy08@gmail.com (J. Peng), DKL5@psu.edu (D.K.J. Lin).

https://doi.org/10.1016/j.jspi.2019.04.011
0378-3758/

Table 1
Designs A and B for 12 runs and 7 factors.

(a) Design A							(b) Design B						
−	−	−	+	+	+	−	−	−	−	−	−	−	−
+	−	−	−	−	+	+	+	−	+	−	−	−	+
−	+	−	−	+	−	+	+	+	−	+	−	−	−
+	+	−	+	−	−	−	−	+	+	−	+	−	−
−	−	+	+	−	−	+	+	−	+	+	−	+	−
+	−	+	−	+	−	−	+	+	−	+	+	−	+
−	+	+	−	−	+	−	+	+	+	−	+	+	−
+	+	+	+	+	+	+	−	+	+	+	−	+	+
0	0	0	0	0	0	0	−	−	+	+	+	−	+
0	0	0	0	0	0	0	−	−	−	+	+	+	−
0	0	0	0	0	0	0	+	−	−	−	+	+	+
0	0	0	0	0	0	0	−	+	−	−	−	+	+

evaluated for both Designs A and B. It is shown that Design B achieves a substantially less dispersed $\hat{\sigma}^2$ than Design A. Section 3 presents the summary of a lengthy simulation study. The simulation results support our theory, and show that the significance tests achieve more reliable performance under Design B. Section 4 is devoted to a slightly different scenario, where designs are restricted to be a (saturated) main-effect design supplemented by several follow-up runs. A theory of the optimal follow-up design is established, and a convenient method is proposed to construct (nearly-)optimal follow-up designs. Concluding remarks are given in Section 5. All proofs are deferred to the Supplementary Material.

2. Main results

The motivating example is a special case of the setup below. Consider, in general, the linear model

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \tag{1}$$

where ϵ_i's are assumed to be i.i.d. with $\mathrm{E}(\epsilon_i) = 0$ and $\mathrm{Var}(\epsilon_i) = \sigma^2$; $\mathbf{X} = (\mathbf{1}, \mathbf{D})$ with $\mathbf{1}$ being the intercept column and $\mathbf{D}$ being an $n \times k$ design matrix ($n > k$). The goal here is to estimate both $\boldsymbol{\beta}$ and σ^2 in an efficient manner. Two types of designs above are considered.

Design A: $\mathbf{D} = \begin{bmatrix} \mathbf{H}_k \\ \mathbf{O} \end{bmatrix}$, where $\mathbf{H}_k$ is any $(k+1) \times k$ H-design and $\mathbf{O}$ is a $(n-k-1) \times k$ zero-matrix. Note that this design is conventionally used in industry.

Design B: $\mathbf{D}$ consists of any k distinct columns of any n-run H-design.

The estimators of $\boldsymbol{\beta}$ and σ^2 can be obtained by the classical least square methods (Seber and Lee, 2012), namely,

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{y}, \text{ and} \tag{2}$$

$$\hat{\sigma}^2 = \frac{1}{n-p}\mathbf{y}^\top(\mathbf{I}_n - \mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top)\mathbf{y}, \tag{3}$$

where $p = k+1$ and $\mathbf{I}_n$ denotes the $n \times n$ identity matrix. It is well known that under the aforementioned model assumptions, $\hat{\boldsymbol{\beta}}$ is always unbiased (under any design), and $\mathrm{Var}(\hat{\boldsymbol{\beta}}) = \sigma^2\left(\mathbf{X}^\top\mathbf{X}\right)^{-1}$ (Seber and Lee, 2012). Note that $\mathbf{X}^\top\mathbf{X}$ is equal to $\begin{bmatrix} n & \mathbf{O} \\ \mathbf{O} & p\mathbf{I}_k \end{bmatrix}$ and $n\mathbf{I}_p$, for Designs A and B respectively. Thus Design B attains a slightly smaller $\mathrm{Var}(\hat{\beta}_j)$ for any $1 \le j \le k$. That is, Design B is (slightly) more efficient in estimating all main effects.

For the estimation of σ^2, it is known that $\hat{\sigma}^2$ follows $\chi^2_{n-p} \cdot \sigma^2/(n-p)$, if ϵ_i's follow a normal distribution (Anderson, 1958). Here we study how the choice of design will affect (i) the variance of $\hat{\sigma}^2$ and (ii) the covariance between $\hat{\sigma}^2$ and each $\hat{\beta}_j$, under various error distributions. The following result (Bai and Silverstein, 2010) is useful to our work. Recall that $p = k+1$, $\mathbf{I}_n$ is the $n \times n$ identity matrix, and $\mathbf{X} = [\mathbf{1}, \mathbf{D}]$ is the model matrix. For the $\hat{\sigma}^2$ given by Eq. (3), it is shown that:

Lemma 1 (*Bai and Silverstein, 2010*).

$$\mathrm{Var}(\hat{\sigma}^2) = \sigma^4\left(\frac{2}{n-p} + \frac{\mathrm{E}(\epsilon_1^4)/\sigma^4 - 3}{(n-p)^2} \cdot \sum_{i=1}^{n} G_{ii}^2\right), \tag{4}$$

where $(G_{11}, G_{22}, \ldots, G_{nn})$ are the diagonal elements of the matrix $\mathbf{I}_n - \mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top$.

Table 2
Theoretical values of $\text{Var}(\hat{\sigma}^2)/\sigma^4$ for Designs A and B under different error distributions and design sizes[a].

Distribution[b]	Excess kurtosis	Design size (n, p)		
		(12, 8)	(20, 16)	(40, 32)
Normal	0	0.5 / 0.5	0.5 / 0.5	0.25 / 0.25
Laplace	3	1.13 / 0.75	1.18 / 0.65	0.607 / 0.325
Exponential	6	1.77 / 1	1.85 / 0.8	0.963 / 0.4
$\chi^2(3)$	4	1.34 / 0.83	1.40 / 0.7	0.725 / 0.35
$t(5)$	6	1.77 / 1	1.85 / 0.8	0.963 / 0.4
Log-normal$(0, 0.5^2)$	5.90	1.74 / 0.992	1.83 / 0.795	0.951 / 0.397
Log-normal$(0, 1)$	111	23.9 / 9.74	25.5 / 6.05	13.4 / 3.02
GEV (shape= 0)	2.4	1.01 / 0.7	1.04 / 0.62	0.535 / 0.31
Uniform	−1.2	0.247 / 0.4	0.229 / 0.44	0.107 / 0.22
Beta(2,2)	−0.857	0.319 / 0.429	0.306 / 0.457	0.148 / 0.229
$0.5N(0, 1) + 0.5N(0, 2^2)$	1.08	0.728 / 0.59	0.744 / 0.554	0.378 / 0.277
$0.5N(-1, 1) + 0.5N(1, 1)$	−0.5	0.395 / 0.458	0.387 / 0.475	0.191 / 0.238
$0.6N(0, 1) + 0.2N(-1, 2^2) + 0.2N(1, 2^2)$	0.83	0.835 / 0.632	0.858 / 0.579	0.438 / 0.290

[a]In the third, fourth, and fifth columns of table, the first number is the value of $\text{Var}(\hat{\sigma}^2)/\sigma^4$ for Design A, and the second number is for Design B.
[b]Every error distribution is linearly transformed so that $E(\epsilon_i) = 0$. Note that the kurtosis does not change under linear transformations.

The error distribution affects $\text{Var}(\hat{\sigma}^2)$ through the quantity $\text{E}(\epsilon_1^4)/\sigma^4 - 3$, the so-called excess kurtosis (EK) of ϵ_i's. For normal ϵ_i's, the excess kurtosis equals 0, and thus $\text{Var}(\hat{\sigma}^2) = 2\sigma^4/(n-p)$. For distributions with a positive EK, $\text{Var}(\hat{\sigma}^2)$ is greater than $2\sigma^4/(n-p)$; while for negative-EK distributions, $\text{Var}(\hat{\sigma}^2) < 2\sigma^4/(n-p)$.

The choice of design affects $\text{Var}(\hat{\sigma}^2)$ through the quantity $\sum_{i=1}^n G_{ii}^2$. When the error distribution has EK>0, from Eq. (4), a minimum $\sum_{i=1}^n G_{ii}^2$ yields a minimum $\text{Var}(\hat{\sigma}^2)$. Specifically, as $\sum_{i=1}^n G_{ii}^2$ increases by 1, $\text{Var}(\hat{\sigma}^2)$ increases by the amount of $\sigma^4 \cdot \text{EK}\ /(n-p)^2$. Note that EK has no upper bound and could be quite large (see Table 2 for examples). As such, $\sum_{i=1}^n G_{ii}^2$ plays a critical rule for the purpose of reducing $\text{Var}(\hat{\sigma}^2)$. On the other hand, when EK<0, $\text{Var}(\hat{\sigma}^2)$ is below $2\sigma^4/(n-p)^2$ (thus is reasonably small), and EK has a lower bound of −2. In such case, $\sum_{i=1}^n G_{ii}^2$ is not so critical. Since the error distribution is typically unknown in practice, we recommend, in general, a design with minimum $\sum_{i=1}^n G_{ii}^2 -$ thus Design B. This is formulated in Theorem 1 .

Theorem 1. *Define G_{ii} as the ith diagonal element of $\mathbf{I}_n - \mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top$.*

(i) For Design A, $G_{ii} = 1/p - 1/n$ $(1 \le i \le p)$, and $G_{ii} = 1 - 1/n$ $(p+1 \le i \le n)$. Therefore $\sum_{i=1}^n G_{ii}^2 = (n-p)(np-2p+1)/(np)$.

(i) For Design B, $G_{ii} = 1 - p/n$ $(1 \le i \le n)$. Therefore $\sum_{i=1}^n G_{ii}^2 = (n-p)^2/n$.

(iii) For any $n \times p$ matrix $\mathbf{X}$, $\sum_{i=1}^n G_{ii}^2 \ge (n-p)^2/n$.

The values of $\text{Var}(\hat{\sigma}^2)$ under different scenarios can thus be theoretically derived, as shown in Table 2. These values provide intuitive comparisons between Designs A and B in the efficiency of estimating σ^2. Note that the values of $\sum_{i=1}^n G_{ii}^2$ in Theorem 1 do not depend on the choice of H-design in Designs A and B. Nor do the values in Table 2.

We next evaluate $\text{Cov}(\hat{\sigma}^2, \hat{\beta}_j)$'s under different designs. As will be shown below, $\hat{\sigma}^2$ is uncorrelated with every $\hat{\beta}_j$ $(j \ge 1)$ under both Designs A and B.

Lemma 2.

$$\text{Cov}(\hat{\boldsymbol{\beta}}, \hat{\sigma}^2) = \frac{\text{E}\,\epsilon_1^3}{n-p}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top[G_{11}, G_{22}, \ldots, G_{pp}]^\top, \tag{5}$$

where $\text{Cov}(\hat{\boldsymbol{\beta}}, \hat{\sigma}^2) = \left[\text{Cov}(\hat{\sigma}^2, \hat{\beta}_0), \text{Cov}(\hat{\sigma}^2, \hat{\beta}_1), \ldots, \text{Cov}(\hat{\sigma}^2, \hat{\beta}_k)\right]^\top$, and G_{ii} is the ith diagonal element of $\mathbf{I}_n - \mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top$.

With the G_{ii}'s for Designs A and B given in Theorem 1, it is easy to verify that $\mathbf{X}^\top[G_{11}, G_{22}, \ldots, G_{pp}]^\top = [n-p, 0, \ldots, 0]^\top$ for both designs. On the other hand, $(\mathbf{X}^\top\mathbf{X})^{-1}$ equals $\begin{bmatrix} n^{-1} & \mathbf{0} \\ \mathbf{0} & p^{-1}\mathbf{I}_{p-1} \end{bmatrix}$ for Design A and $n^{-1}\mathbf{I}_p$ for Design B. It follows that $\text{Cov}(\hat{\boldsymbol{\beta}}, \hat{\sigma^2}) = \text{E}\,\epsilon_1^3\,[(n-p)/n, 0, \ldots, 0]^\top$ for both designs. As a result, the estimation of σ^2 is uncorrelated with all the main effects under either Design A or B.

3. Simulation study

Some simulation results are provided here to demonstrate our theories in Section 2. Two examples with different design sizes are given. Section 3.1 studies the case with $(n, p) = (12, 8)$ and Section 3.2 studies $(n, p) = (40, 32)$. Under

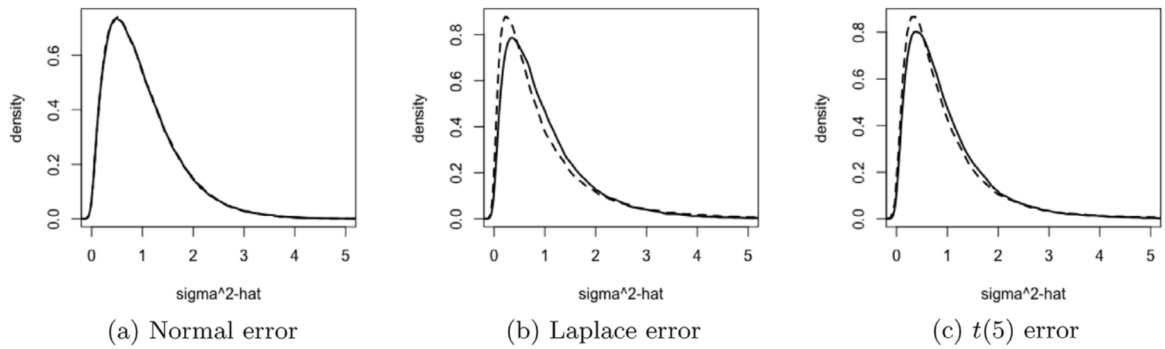

Fig. 1. Sampling distributions of $\hat{\sigma}^2$ under Designs A and B for different error distributions, when $(n, p) = (12, 8)$ (dashed curve: Design A; solid curve: Design B).

Table 3
Quantiles of the sampling distribution of $\hat{\sigma}^2$, when $(n, p) = (12, 8)$.

(a) When the error distribution is Laplace							
Percentage	5%	10%	20%	50%	80%	90%	95%
Quantile under Design A	0.103	0.164	0.276	0.678	1.498	2.205	2.991
Quantile under Design B	0.142	0.216	0.346	0.761	1.510	2.072	2.660
(b) When the error distribution is $t(5)$							
Percentage	5%	10%	20%	50%	80%	90%	95%
Quantile under Design A	0.129	0.197	0.315	0.692	1.419	2.042	2.805
Quantile under Design B	0.150	0.226	0.355	0.758	1.459	2.008	2.610

two common distributions: Laplace and $t(5)$, the numerical results support our conclusion in Section 2 that Design B provides more efficient estimates of σ^2 than Design A. In addition, it is found that Design B, as compared with Design A, achieves a more reliable type-I error and a higher power for the t- and F-tests.

3.1. Case-I: $(n, p) = (12, 8)$

Under different error distributions, we evaluate the sampling distributions of both Designs A and B (with 12 runs and 7 factors). The sampling distribution of $\hat{\sigma}^2$ is obtained via the following way. Set the true $\beta_0 = 2$, $\beta_j = 0.5$ $(1 \le j \le 7)$, and $\sigma^2 = 1$. Each time, draw a sample of ϵ_i's from the given distribution, generate y_i's under the given design, and then estimate σ^2 using Eq. (3). Run this procedure for 100,000 times. The sampling distributions of $\hat{\sigma}^2$ under the given design and error distribution can thus be obtained.

Fig. 1 shows the kernel-smoothed densities (Wand and Jones, 1994) for the sampling distributions of $\hat{\sigma}^2$, under both Designs A and B and different error distributions. The result for each distribution is displayed in Fig. 1, where the density for Design B is represented by a solid curve, and the density for Design A is represented by a dashed curve.

Fig. 1(a) shows the distributions of $\hat{\sigma}^2$ when each ϵ_i follows a standard normal distribution. As expected, the distributions of $\hat{\sigma}^2$ are the same under Designs A and B. Fig. 1(b) shows the distributions of $\hat{\sigma}^2$ when ϵ_i follows Laplace$(0, 1)/\sqrt{2}$ (the divisor $\sqrt{2}$ is for normalizing ϵ_i so that $\sigma^2 = 1$). It is evident $\hat{\sigma}^2$ is less dispersed under Design B, which agrees with Theorem 1. (Under Design A, $\hat{\sigma}^2$ is more skewed towards the right.) As another way of comparison, the quantiles of $\hat{\sigma}^2$ under Designs A and B are displayed in Table 3(a). Almost all the quantiles are closer to the true value $(\sigma^2 = 1)$ under Design B, which clearly indicates that Design B achieves a less dispersed estimate of error variance.

Fig. 1(c) and Table 3(b) exhibit the distributions of $\hat{\sigma}^2$ when ϵ_i follows a $t(5)/\sqrt{3}$ distribution (the divisor $\sqrt{3}$ is chosen to guarantee $\sigma^2 = 1$). The pattern here is very similar to that under Laplace distribution. Both the density plots and quantiles indicate that $\hat{\sigma}^2$ is slightly less dispersed under Design B.

The difference in the sampling distributions of $\hat{\sigma}^2$ under Designs A and B generally becomes more evident when n and p increase, as will be shown below (Section 3.2).

3.2. Case-II: $(n, p) = (40, 32)$

Consider Designs A and B of a different size: $(n, p) = (40, 32)$. The sampling distributions of $\hat{\sigma}^2$ are demonstrated in Fig. 2 and Table 4, which are analogies of those in Case I. We follow the simulation settings in Section 3.1: 100,000 random samples were drawn to generate each plot and table. The true σ^2 is set as 1 and $\boldsymbol{\beta}$ is set as $(2, 0.5, \ldots, 0.5)$.

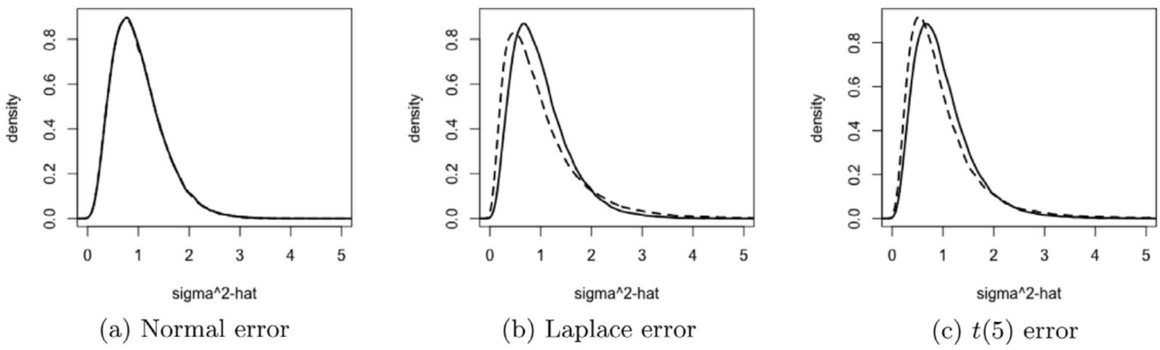

Fig. 2. Sampling distributions of $\hat{\sigma}^2$ under Designs A and B for different error distributions, when $(n, p) = (40, 32)$ (dashed curve: Design A; solid curve: Design B).

Table 4
Quantiles of the sampling distribution of $\hat{\sigma}^2$, when $(n, p) = (40, 32)$.

(a) When the error distribution is Laplace							
Percentage	5%	10%	20%	50%	80%	90%	95%
Quantile under Design A	0.212	0.290	0.414	0.794	1.450	1.960	2.480
Quantile under Design B	0.307	0.397	0.531	0.886	1.401	1.748	2.084
(b) When the error distribution is $t(5)$							
Percentage	5%	10%	20%	50%	80%	90%	95%
Quantile under Design A	0.249	0.326	0.446	0.784	1.365	1.825	2.356
Quantile under Design B	0.306	0.397	0.528	0.876	1.383	1.740	2.096

The patterns here are very similar to those in Section 3.1 (with perhaps slightly more evidence). Under either Laplace or $t(5)$ errors, the sampling distribution of $\hat{\sigma}^2$ is apparently more concentrated towards the true value 1 under Design B than under Design A. (In fact, under Design B, the sampling distribution of $\hat{\sigma}^2$ deviates less than that under normality.)

In many situations, the reliability of significant tests, such as t- and F-tests, are of interest. Intuitively, a more efficient estimate of σ^2 will yield more reliable test results. To confirm this, we next evaluate the powers of t- and F-tests under both Designs A and B. Set the significance level at 0.05. Let β_j's $(1 \le j \le p-1)$ gradually deviate from 0 while fixing the intercept $\beta_0 = 2$. For each true $\boldsymbol{\beta}$, draw 100,000 random samples of $\mathbf{y}$ from model (1) under both Designs A and B. Determine whether each sample of $\mathbf{y}$ falls in the critical regions of (i) the t-test for $H_0 : \beta_1 = 0$ and (ii) the F-test for $H_0 : \beta_j = 0$ (for all $1 \le j \le p-1$), and then estimate the powers of t- and F-tests for this particular $\boldsymbol{\beta}$. Simulations have been conducted, for $(n, p) = (12, 8)$ and $(40, 32)$ respectively, under different settings of $\boldsymbol{\beta}$. The results are summarized in Appendix.

The simulations show that under $t(5)$ distribution (positive EK), the type-I error is always controlled by the significance level under Design B, while the type-I error considerably exceeds the significance level under Design A (especially for the F-test). Moreover, as long as $\boldsymbol{\beta}$ is not too small, Design B achieves a higher power than Design A. It is also shown that results under uniform error distribution (negative EK), the type-I error is well-controlled under both Design A and Design B. On the other hand, Design B still achieves a higher power than Design A under uniform errors. As pointed out by one referee of our work, the higher power under Design B is partially ascribed to its higher D-criterion. It will be explored in the future how the advantages of Design B in significance tests are connected to the fact that Design B attains the minimum $\sum_{i=1}^{n} G_{ii}^2$.

In summary, this section compares the sampling distributions of $\hat{\sigma}^2$ for Designs A and B, under different scenarios. The numerical results demonstrate that Design B achieves a less dispersed $\hat{\sigma}^2$; this supports the theory in Section 2. It is also shown that Design B achieves robust type-I errors (i.e., the type-I errors are closer to the significance level) and higher powers of t- and F-tests, as compared to Design A.

4. Optimal follow-up design

As a sequel, we study the design problem following an initial main-effect design. Suppose an experimenter has first conducted a saturated main-effect experiment with an $p \times k$ design, $\mathbf{D}_0$ (recall that $p = k + 1$). It is well known that under the effect sparsity assumption (Box and Meyer, 1986), a saturated main-effect design is useful for identifying the significant factors. Many inference methods are available under effect sparsity, for example, Daniel (1959), Dong (1993), Lenth (1989), Ye et al. (2001), and Miller (2005). However, the effect sparsity assumption may not hold in some situations (see, e.g., Hurley, 1995). In such cases, a follow-up experiment is needed to provide extra degrees of freedom for estimating error variance.

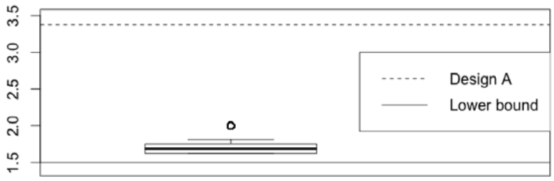

Fig. 3. The distribution of the quantity $\sum_{i=1}^{n} G_{ii}^2$ for the obtained designs.

Our goal is to find an optimal $(n-p)\times k$ design $\mathbf{D}_a$, so that the combined design $\mathbf{D} = \begin{bmatrix}\mathbf{D}_0\\ \mathbf{D}_a\end{bmatrix}$ achieves an efficient estimate of both $\boldsymbol{\beta}$ and σ^2. Again, $\mathbf{D}$ is slightly larger than $\mathbf{D}_0$, i.e., n exceeds p by only a few runs. Assume that $\mathbf{D}_0$ is an H-design, and each element of $\mathbf{D}_a$ is within the interval $[-1, 1]$. In the following discussion, denote $\mathbf{X} = (\mathbf{1}, \mathbf{D})$, $\mathbf{X}_0 = (\mathbf{1}, \mathbf{D}_0)$, and $\mathbf{X}_a = (\mathbf{1}, \mathbf{D}_a)$, where $\mathbf{1}$ is the intercept column.

Remark 1. The conventional choice of $\mathbf{D}_a$ is $(n-p)$ runs of center points (so that Design A is the combined design). Design B is not considered in this section, because it does not include any $p \times k$ orthogonal sub-design.

In essence, we have three criteria for finding an optimal $\mathbf{D}_a$: (A) maximize the efficiency of $\hat{\boldsymbol{\beta}}$, (B) minimize the quantity $\sum_{i=1}^{n} G_{ii}^2$ (see the discussion in Section 2), and (C) minimize the absolute value of $\mathrm{Cov}(\hat{\sigma}^2, \hat{\beta}_j)$ for each j.

For item (A), the estimation of $\boldsymbol{\beta}$, the variance–covariance matrix of $\hat{\boldsymbol{\beta}}$ is proportional to $\left(\mathbf{X}^\top\mathbf{X}/n\right)^{-1}$. Thus it is desirable to maximize the moment matrix $\mathbf{X}^\top\mathbf{X}/n$ in terms of, say, the conventional D-criterion:

$$\arg\max_{\mathbf{D}_a} \left|\mathbf{X}^\top\mathbf{X}/n\right|^{1/p},$$

where $|\cdot|$ denotes the determinant of matrix. Note that $\mathbf{X}^\top\mathbf{X} = \mathbf{X}_0^\top\mathbf{X}_0 + \mathbf{X}_a^\top\mathbf{X}_a = p\mathbf{I}_p + \mathbf{X}_a^\top\mathbf{X}_a$. By the Sylvester's determinant identity, $\left|\mathbf{X}^\top\mathbf{X}/p\right| = \left|\mathbf{I}_{n-p} + \mathbf{X}_a\mathbf{X}_a^\top/p\right|$. As each element of $\mathbf{X}_a$ is within the interval $[-1, 1]$, it is easy to see that $\mathbf{X}_a\mathbf{X}_a^\top/p \preceq \mathbf{I}_{n-p}$ in the Loewner order (i.e., $\mathbf{I}_{n-p} - \mathbf{X}_a\mathbf{X}_a^\top/p$ is positive semi-definite). It follows that

Theorem 2. *For the follow-up design problem, the combined design achieves the D-optimum if and only if $\mathbf{X}_a\mathbf{X}_a^\top = p\mathbf{I}_{n-p}$. Such D-optimal combined design has a D-criterion of*

$$\left|\mathbf{X}^\top\mathbf{X}/n\right|^{1/p} = \frac{p}{n} \cdot 2^{(n-p)/p}.$$

Note that $\mathbf{X}_a\mathbf{X}_a^\top = p\mathbf{I}_{n-p}$ holds if and only if (i) each element of $\mathbf{D}_a$ equals 1 or -1 and (ii) the rows of $\mathbf{X}_a$ are orthogonal to each other. In particular, the equality holds when $\mathbf{D}_a$ is a fraction of any H-design.

For item (B), the purpose of minimizing the quantity $\sum_{i=1}^{n} G_{ii}^2$, we derive a lower bound of this quantity among all possible follow-up designs:

Theorem 3. *For any choice of $\mathbf{D}_a$, it holds that*

$$\sum_{i=1}^{n} G_{ii}^2 \geq \frac{(n-p)n}{4p}. \tag{6}$$

In general, such lower bound may not be attainable. A convenient method is proposed here of which the resulting designs achieve or nearly achieve the lower bound in $\sum_{i=1}^{n} G_{ii}^2$, and are D-optimal in estimating $\hat{\boldsymbol{\beta}}$. First, obtain a design $\mathbf{D}_1$ which consists of any $n-p$ distinct rows of $\mathbf{D}_0$. Second, obtain the desired $\mathbf{D}_a$ by randomly rearranging the factor indices, i.e., the columns of $\mathbf{D}_1$.

By Theorem 2, the resulting $\mathbf{D}_a$'s are D-optimal. They are also highly efficient in estimating σ^2, and here is an example. Consider the previous case where $n = 12$ and $k = 7$ (and thus $p = 8$). 10,000 designs are generated using the above method. For the obtained designs, the distribution of $\sum_{i=1}^{n} G_{ii}^2$ is shown as the boxplot in Fig. 1. The solid horizontal line in the graph indicates the lower bound of $\sum_{i=1}^{n} G_{ii}^2$ (see Theorem 2), which is 1.5. The dashed line indicates the quantity under Design A, which is 3.375. It is evident that the obtained designs are nearly optimal, and are much more preferable than Design A in terms of minimizing $\sum_{i=1}^{n} G_{ii}^2$ (see Fig. 3).

For item (C), the criterion of minimizing $\left|\mathrm{Cov}(\hat{\sigma}^2, \hat{\beta}_j)\right|$, this criterion only matters when the error distribution is skewed. (For symmetric distributions, $\mathrm{E}(\epsilon_i^3) = 0$, and according to Lemma 2, $\mathrm{Cov}(\hat{\sigma}^2, \hat{\beta}_j)$ is always 0 for any choice of design.) Under skewed error distributions, Design A still achieves $\mathrm{Cov}(\hat{\sigma}^2, \hat{\beta}_j) = 0$ for each $1 \leq j \leq k$ (see Section 2). As for follow-up designs obtained by the above algorithm, $\mathrm{Cov}(\hat{\sigma}^2, \hat{\beta}_j)$'s are no longer equal to 0. However, our empirical study indicates that the correlations are typically very small.

In summary, this section discusses how to choose a follow-up design for an H-design so that the combined design is (A) D-optimal, (B) achieves minimum $\sum_{i=1}^{n} G_{ii}^2$, or (C) achieves zero-correlation between $\hat{\sigma}^2$ and all main effects. Optimality theories are given regarding the three criteria respectively. We have proposed a convenient construction of designs that are optimal in terms of criterion (A) and nearly optimal in terms of criteria (B) and (C).

5. Concluding remarks

This article studies the choice of an $n \times k$ design for estimating both factor effects ($\boldsymbol{\beta}$) and error variance (σ^2) via a linear main-effect model, when the number of runs (n) is slightly greater than the number of factors (k). The main focus is to find the optimal design for best estimating σ^2. The conventional Design A (an H-design appended by center points) and a proposed Design B (the projection of a larger H-design) are under study. The main conclusion is that compared with Design A, Design B typically achieves a more reliable estimate of σ^2, denoted by $\hat{\sigma}^2$. We demonstrate this conclusion via an optimality theory on the criterion Var($\hat{\sigma}^2$), as well as simulation studies to compare the sampling distributions of $\hat{\sigma}^2$ for Designs A and B. More simulation results indicate that the t- and F-tests are more reliable (in terms of power and type-I error) under Design B, as compared to Design A. We also consider a practical scenario where the design is constrained to be a follow-up of a main-effect saturated design. Theories on the follow-up designs are established, and a convenient algorithm has been proposed to generate follow-up designs which are optimal in estimating $\boldsymbol{\beta}$ and (nearly-)optimal in estimating σ^2.

There are some limitations of our work. This paper assumes that n and p are both multipliers of 4. Future work needs to be done for more flexible design sizes. In addition, note that when n is much greater than p, different designs will not vary much in the efficiency of estimating σ^2: from Lemma 1, the effect of $\sum_{i=1}^{n} G_{ii}^2$ on Var($\hat{\sigma}^2$) is shrunk by the multiplier $(n-p)^{-2}$. Our results are only useful when $n-p$ is small compared to n. Furthermore, the mean-squared error (MSE) is used here as the estimator of σ^2. If the model is inadequate (such as when some important interaction effects are excluded in the model) MSE will be a biased estimator of the pure error variance. In this case, it is recommended to estimate σ^2 with *partial replications*, so that the lack-of-fit is separated from pure error. See, for example, Gilmour and Trinca (2012), Tsai and Liao (2014), Jones and Montgomery (2017), and Leonard and Edwards (2017).

When the model is adequate (the lack-of-fit is small relative to the pure error variance) and a small $n-p$ is desirable, our work provides useful guidance for practitioners to properly choose a design, so as to efficiently estimate error variance and identify the active factors. It is anticipated that some of the proposed techniques here, e.g., the methods of proving the optimality theory, will be useful for a wider class of models (such as those models with interaction or higher-order effects).

It is notable that there can be more than one Hadamard matrices given the design size, so that the choice of Designs A and B is generally not unique. While it does not affect the theoretical properties of Designs A and B under the assumptions in Section 2, the choice of underlying Hadamard matrices does matter in the presence of nuisance interaction effects. In terms of this, it is recommended to use a Plackett–Burman design for the underlying Hadamard matrix of Design B, as explained below. Lin and Draper (1992, 1993) showed that in a Plackett–Burman design, the absolute correlations between the two-interactions and the main effects are typically small. For example, with $n = 12$, the maximum absolute correlation between two-way interactions and main effects is only 1/3 (Lin and Draper 1993). Thus, if Design B is obtained by choosing partial columns of a Plackett–Burman design, the nuisance (two-way) interaction effects will not substantially bias the main effects. On the other hand, in Design A, the interactions and the main effects will have higher absolution correlations because of the center points. Thus, the main effects estimates are typically more biased under Design A than under Design B. As for the statistical power in the presence of nuisance interaction effects, note that Design B has a higher D-criterion while it inflates the estimation of error, on the contrary. In view of this, whether Design B or A is more powerful depends on the specific design size and magnitude of nuisance effects.

Acknowledgments

The authors thank Bradley Jones for bringing their attention into this problem, and also the reviewers for their constructive comments.

Appendix. More simulation results

Tables A.1 and A.2 compares the performance of Designs A and B, in terms of the t- and F-tests. See Section 3.2 for how the tables were obtained.

Appendix B. Supplementary material: proofs

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.jspi.2019.04.011.

Table A.1
Powers of t- and F-tests under different β's for Designs A and B, when $(n, p) = (12, 8)$.

(a) When the error distribution is $t(5)$

True β	t-test for β_1		F-test	
	Under Design A	Under Design B	Under Design A	Under Design B
$\beta_j = 0$ $(1 \le j \le 31)$	0.0656	0.0457	0.0799	0.0468
$\beta_j = 0.8$ $(1 \le j \le 31)$	0.2423	0.2991	0.334	0.4108
$\beta_j = 1.6$ $(1 \le j \le 31)$	0.621	0.7609	0.7932	0.9016
$\beta_j = 2.4$ $(1 \le j \le 31)$	0.8824	0.9555	0.9574	0.9904
$\beta_j = 3.2$ $(1 \le j \le 31)$	0.9679	0.9916	0.9902	0.9987
$\beta_1 = 0.5$, $\beta_j = 0$ $(2 \le j \le 31)$	0.4665	0.5944	0.1664	0.1702
$\beta_1 = 1$, $\beta_j = 0$ $(2 \le j \le 31)$	0.9085	0.9676	0.4473	0.5556
$\beta_1 = 1.5$, $\beta_j = 0$ $(2 \le j \le 31)$	0.9874	0.9981	0.7283	0.8539
$\beta_1 = 2$, $\beta_j = 0$ $(2 \le j \le 31)$	0.9977	0.9998	0.8934	0.9627

(b) When the error distribution is Uniform

True β	t-test for β_1		F-test	
	Under Design A	Under Design B	Under Design A	Under Design B
$\beta_j = 0$ $(1 \le j \le 31)$	0.0320	0.0542	0.0191	0.0554
$\beta_j = 0.5$ $(1 \le j \le 31)$	0.1543	0.2528	0.1763	0.3496
$\beta_j = 1$ $(1 \le j \le 31)$	0.5429	0.724	0.7797	0.9201
$\beta_j = 1.5$ $(1 \le j \le 31)$	0.898	0.9776	0.9977	0.9998
$\beta_j = 2$ $(1 \le j \le 31)$	0.9943	1	1	1
$\beta_1 = 0.8$, $\beta_j = 0$ $(2 \le j \le 31)$	0.3636	0.5324	0.064	0.1602
$\beta_1 = 1.6$, $\beta_j = 0$ $(2 \le j \le 31)$	0.9348	0.9908	0.2772	0.4876
$\beta_1 = 2.4$, $\beta_j = 0$ $(2 \le j \le 31)$	0.9998	1	0.6723	0.85
$\beta_1 = 3.2$, $\beta_j = 0$ $(2 \le j \le 31)$	1	1	0.9472	0.9879

Table A.2
Powers of t- and F-tests under different β's for Designs A and B, when $(n, p) = (40, 32)$.

(a) When the error distribution is $t(5)$

True β	t-test for β_1		F-test	
	Under Design A	Under Design B	Under Design A	Under Design B
$\beta_j = 0$ $(1 \le j \le 31)$	0.0784	0.0475	0.1333	0.0468
$\beta_j = 0.2$ $(1 \le j \le 31)$	0.2141	0.2114	0.4507	0.4518
$\beta_j = 0.4$ $(1 \le j \le 31)$	0.5461	0.62	0.8976	0.9733
$\beta_j = 0.6$ $(1 \le j \le 31)$	0.8362	0.907	0.992	0.9999
$\beta_j = 0.8$ $(1 \le j \le 31)$	0.9611	0.9876	0.9994	1
$\beta_1 = 0.4$, $\beta_j = 0$ $(2 \le j \le 31)$	0.5457	0.6206	0.1736	0.0832
$\beta_1 = 0.8$, $\beta_j = 0$ $(2 \le j \le 31)$	0.9628	0.9876	0.3001	0.2338
$\beta_1 = 1.2$, $\beta_j = 0$ $(2 \le j \le 31)$	0.9992	1	0.4961	0.5199
$\beta_1 = 1.6$, $\beta_j = 0$ $(2 \le j \le 31)$	1	1	0.6941	0.7912

(b) When the error distribution is Uniform

True β	t-test for β_1		F-test	
	Under Design A	Under Design B	Under Design A	Under Design B
$\beta_j = 0$ $(1 \le j \le 31)$	0.0348	0.0501	0.0116	0.0512
$\beta_j = 0.2$ $(1 \le j \le 31)$	0.1439	0.196	0.242	0.4234
$\beta_j = 0.4$ $(1 \le j \le 31)$	0.4948	0.5968	0.9891	0.9854
$\beta_j = 0.6$ $(1 \le j \le 31)$	0.8529	0.9158	1	1
$\beta_j = 0.8$ $(1 \le j \le 31)$	0.9829	0.9944	1	1
$\beta_1 = 0.4$, $\beta_j = 0$ $(2 \le j \le 31)$	0.4948	0.5973	0.022	0.0841
$\beta_1 = 0.8$, $\beta_j = 0$ $(2 \le j \le 31)$	0.9827	0.9947	0.0858	0.2194
$\beta_1 = 1.2$, $\beta_j = 0$ $(2 \le j \le 31)$	1	1	0.3072	0.49
$\beta_1 = 1.6$, $\beta_j = 0$ $(2 \le j \le 31)$	1	1	0.6932	0.7886

References

Anderson, T.W., 1958. An Introduction to Multivariate Statistical Analysis, Vol. 2. Wiley New York.
Bai, Z., Silverstein, J.W., 2010. Spectral Analysis of Large Dimensional Random Matrices, Vol. 20. Springer.
Box, G.E.P., Meyer, R.D., 1986. An analysis for unreplicated fractional factorials. Technometrics 28 (1), 11–18.
Daniel, C., 1959. Use of half-normal plots in interpreting factorial two-level experiments. Technometrics 1 (4), 311–341.
Dong, F., 1993. On the identification of active contrasts in unreplicated fractional factorials. Statist. Sinica 209–217.
Gilmour, S.G., Trinca, L.A., 2012. Optimum design of experiments for statistical inference. J. R. Stat. Soc. Ser. C. Appl. Stat. 61 (3), 345–401.
Hurley, P.D., 1995. The conservative nature of the effect sparsity assumption for saturated fractional factorial experiments. Qual. Eng. 7 (4), 657–671.

Jones, B., Montgomery, D.C., 2017. Partial replication of small two-level factorial designs. Qual. Eng. 29 (2), 190–195.
Lenth, R.V., 1989. Quick and easy analysis of unreplicated factorials. Technometrics 31 (4), 469–473.
Leonard, R.D., Edwards, D.J., 2017. Bayesian D-optimal screening experiments with partial replication. Comput. Statist. Data Anal..
Lin, D.K., Draper, N.R., 1992. Projection properties of Plackett and Burrnan designs. Technometrics 34 (4), 423–428.
Lin, D.K., Draper, N.R., 1993. Generating alias relationships for two-level Plackett and Burman designs. Comput. Stat. Data Anal. 15 (2), 147–157.
Miller, A., 2005. The analysis of unreplicated factorial experiments using all possible comparisons. Technometrics 47 (1), 51–63.
Plackett, R.L., Burman, J.P., 1946. The design of optimum multifactorial experiments. Biometrika 33 (4), 305–325.
Seber, G.A., Lee, A.J., 2012. Linear Regression Analysis, Vol. 936. John Wiley & Sons.
Tsai, S., Liao, C., 2014. Selection of partial replication on two-level orthogonal arrays. Canad. J. Statist. 42 (1), 168–183.
Wand, M.P., Jones, M.C., 1994. Kernel Smoothing. CRC Press.
Ye, K.Q., Hamada, M., Wu, C.J., 2001. A step-down lenth method for analyzing unreplicated factorial designs. J. Qual. Technol. 33 (2), 140.