Notes: Cluster, don't replicate

Douglas P. Wiens[1]

April 7, 2019

Abstract There is some, not inconsiderable, resistance to the adoption of robust methods –and not only in design. In design this might originate because so many of the 'optimally robust' designs are computationally quite intensive. Here we consider situations in which the classically optimal designs have been obtained, and call for replication at a small number of design points. We put forward the thesis that a fair degree of robustness can be obtained, simply by spreading the replicates out into clusters of nearby, but distinct, locations.

1 Introduction

- Experimenter fits (the 'null' model) a, linear or nonlinear, response

$$E_0\left[Z|\boldsymbol{x}\right] = \eta_0\left(\boldsymbol{x};\boldsymbol{\theta}\right), \tag{1}$$

  with a p-dimensional parameter vector $\boldsymbol{\theta}$, where the independent variables $\boldsymbol{x}$ are chosen from a design space $\chi \subset \mathbb{R}^q$ of finite measure. We denote by $\mu\left(d\boldsymbol{x}\right)$ the (continuous or discrete) uniform probability measure on χ. Unbeknownst to the experimenter, the true ('alternate' model) response is

$$E_1\left[Z|\boldsymbol{x}\right] = \eta_1\left(\boldsymbol{x}\right). \tag{2}$$

- As is common, if the model is truly nonlinear we consider locally optimal designs, in which the experimenter specifies an initial estimate $\boldsymbol{\theta}_*$, expands around it, and fits the resulting linear approximation model. For this, define

$$\boldsymbol{f}\left(\boldsymbol{x}\right) = \dot{\eta}_0\left(\boldsymbol{x};\boldsymbol{\theta}_*\right) \text{ and } Y = Z - \left[\eta_0\left(\boldsymbol{x};\boldsymbol{\theta}_*\right) - \boldsymbol{f}'\left(\boldsymbol{x}\right)\boldsymbol{\theta}_*\right],$$

  obtaining the null approximation

$$E_0\left[Y|\boldsymbol{x}\right] = \boldsymbol{f}'\left(\boldsymbol{x}\right)\boldsymbol{\theta}, \tag{3}$$

  with regressors $\boldsymbol{f}\left(\boldsymbol{x}\right)$.

[1] Department of Mathematical and Statistical Sciences; University of Alberta, Edmonton, Alberta;Canada T6G 2G1. e-mail: `doug.wiens@ualberta.ca`

- Define
$$\boldsymbol{M}_\mu = \int_\chi \boldsymbol{f}\left(\boldsymbol{x}\right)\boldsymbol{f}'\left(\boldsymbol{x}\right)\mu\left(d\boldsymbol{x}\right),$$
and assume that $\boldsymbol{M}_\mu$ is positive definite. We define the 'target' parameter $\boldsymbol{\theta}_0$, effecting the best L_2-approximation of the alternate response by the linear approximation to the null responses, by
$$\boldsymbol{\theta}_0 = \arg\min_{\boldsymbol{\theta}} D_1\left(\boldsymbol{\theta}\right), \tag{4}$$
where
$$\begin{aligned} D_1\left(\boldsymbol{\theta}\right) &= \int_\chi \left\{E_1\left[Z|\boldsymbol{x}\right] - E_0\left[Z|\boldsymbol{x}\right]\right\}^2 \mu\left(d\boldsymbol{x}\right) \\ &= \int_\chi \left\{E_1\left[Z|\boldsymbol{x}\right] - E_0\left[Y + \left(\eta_0\left(\boldsymbol{x};\boldsymbol{\theta}_*\right) - \boldsymbol{f}'\left(\boldsymbol{x}\right)\boldsymbol{\theta}_*\right)\right]\right\}^2 \mu\left(d\boldsymbol{x}\right) \\ &= \int_\chi \left\{\eta_1\left(\boldsymbol{x}\right) - \eta_0\left(\boldsymbol{x};\boldsymbol{\theta}_*\right) - \boldsymbol{f}'\left(\boldsymbol{x}\right)\left(\boldsymbol{\theta} - \boldsymbol{\theta}_*\right)\right\}^2 \mu\left(d\boldsymbol{x}\right). \end{aligned} \tag{5}$$
Define also
$$\tau^2 = D_1\left(\boldsymbol{\theta}_0\right) = \min_{\boldsymbol{\theta}} D_1\left(\boldsymbol{\theta}\right). \tag{6}$$

- **Choice of $\boldsymbol{\theta}_*$.** The 'best' starting value is
$$\boldsymbol{\theta}_{*,0} = \arg\min_{\boldsymbol{\theta}_*} D_0\left(\boldsymbol{\theta}_*\right), \tag{7}$$
where
$$D_0\left(\boldsymbol{\theta}\right) = \int_\chi \left\{\eta_1\left(\boldsymbol{x}\right) - \eta_0\left(\boldsymbol{x};\boldsymbol{\theta}\right)\right\}^2 \mu\left(d\boldsymbol{x}\right).$$
Assuming that $\boldsymbol{\theta}_{*,0}$ is a critical point,
$$\int_\chi \left\{\eta_1\left(\boldsymbol{x}\right) - \eta_0\left(\boldsymbol{x};\boldsymbol{\theta}_{*,0}\right)\right\}\boldsymbol{f}\left(\boldsymbol{x}\right)\mu\left(d\boldsymbol{x}\right) = \boldsymbol{0},$$
and so if this starting value is used, we have $\boldsymbol{\theta}_* = \boldsymbol{\theta}_{*,0}$ and so
$$\begin{aligned} D_1\left(\boldsymbol{\theta}\right) &= \int_\chi \left\{\eta_1\left(\boldsymbol{x}\right) - \eta_0\left(\boldsymbol{x};\boldsymbol{\theta}_{*,0}\right) - \boldsymbol{f}'\left(\boldsymbol{x}\right)\left(\boldsymbol{\theta} - \boldsymbol{\theta}_{*,0}\right)\right\}^2 \mu\left(d\boldsymbol{x}\right) \\ &= D_0\left(\boldsymbol{\theta}_{*,0}\right) + \left(\boldsymbol{\theta} - \boldsymbol{\theta}_{*,0}\right)'\boldsymbol{M}_\mu\left(\boldsymbol{\theta} - \boldsymbol{\theta}_{*,0}\right)' \end{aligned}$$
with a minimum at $\boldsymbol{\theta}_0 = \boldsymbol{\theta}_{*,0}$. We define
$$\tau_0^2 = D_0\left(\boldsymbol{\theta}_{*,0}\right) = \min_{\boldsymbol{\theta}} D_0\left(\boldsymbol{\theta}\right).$$

- Let ξ be a design on χ intended for (3), with I points of support, placing n_i observations at $\boldsymbol{x}_i$ ($\sum_{i=1}^{I} n_i = n$). We also view ξ as a probability measure, in terms of which

$$\begin{aligned}
\boldsymbol{M}_\xi &= \int_\chi \boldsymbol{f}(\boldsymbol{x})\,\boldsymbol{f}'(\boldsymbol{x})\,\xi(d\boldsymbol{x}) = \sum_{i=1}^{I} \frac{n_i}{n}\boldsymbol{f}(\boldsymbol{x}_i)\,\boldsymbol{f}'(\boldsymbol{x}_i),\\
\boldsymbol{r}_\xi &= \int_\chi \boldsymbol{f}(\boldsymbol{x})\,\eta_1(\boldsymbol{x})\,\xi(d\boldsymbol{x}) = \sum_{i=1}^{I} \frac{n_i}{n}\boldsymbol{f}(\boldsymbol{x}_i)\,\eta_1(\boldsymbol{x}_i)\\
\boldsymbol{s}_\xi &= \int_\chi \boldsymbol{f}(\boldsymbol{x})\,\eta_0(\boldsymbol{x};\boldsymbol{\theta}_*)\,\xi(d\boldsymbol{x}) = \sum_{i=1}^{I} \frac{n_i}{n}\boldsymbol{f}(\boldsymbol{x}_i)\,\eta_0(\boldsymbol{x}_i;\boldsymbol{\theta}_*).
\end{aligned}$$

The covariance matrix and bias vector of the least squares estimate $\hat{\boldsymbol{\theta}}$, when the data actually arise from (2) and the error variance is σ^2, are

$$\begin{aligned}
\mathrm{COV}\left[\hat{\boldsymbol{\theta}}\right] &= \sigma^2\left[\boldsymbol{X}'(\xi)\,\boldsymbol{X}(\xi)\right]^{-1} = \frac{\sigma^2}{n}\boldsymbol{M}_\xi^{-1}, && (8)\\
\mathrm{BIAS}\left[\hat{\boldsymbol{\theta}}\right] &= E\left[\hat{\boldsymbol{\theta}}\right] - \boldsymbol{\theta}_0\\
&= \boldsymbol{M}_\xi^{-1}\sum_{i=1}^{I}\frac{n_i}{n}\boldsymbol{f}(\boldsymbol{x}_i)\left\{\eta_1(\boldsymbol{x}_i) - \eta_0(\boldsymbol{x}_i;\boldsymbol{\theta}_*)\right\} - (\boldsymbol{\theta}_0 - \boldsymbol{\theta}_*)\\
&= \boldsymbol{M}_\xi^{-1}(\boldsymbol{r}_\xi - \boldsymbol{s}_\xi) - (\boldsymbol{\theta}_0 - \boldsymbol{\theta}_*). && (9)
\end{aligned}$$

- Let ξ_0 be a design on χ which is optimal for (3), with respect to some optimality criterion $\phi(\xi)$ (typically based on $\boldsymbol{M}_\xi$) to be minimized. Suppose that the study is of size n, and that then the implementation of ξ_0 calls for replicates to be made at several locations. We intend to construct and investigate 'cluster' designs $\tilde{\xi}$ in which the replicates called for by ξ_0 are instead spread out amongst nearby locations. We will compare:

  1. $\phi\left(\tilde{\xi}\right)$ with $\phi(\xi_0)$; of course the latter is smaller but we hope to find that the relative efficiency, as measured by $\phi(\xi_0)/\phi\left(\tilde{\xi}\right)$, is not much less than one.
  2. The behaviour of the two designs under the alternate model, as measured by the integrated mean square error (IMSE) of

     $$\hat{Z}(\boldsymbol{x}) = [\eta_0(\boldsymbol{x};\boldsymbol{\theta}_*) - \boldsymbol{f}'(\boldsymbol{x})\,\boldsymbol{\theta}_*] + \hat{Y}(\boldsymbol{x}) = \eta_0(\boldsymbol{x};\boldsymbol{\theta}_*) + \boldsymbol{f}'(\boldsymbol{x})\left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_*\right),$$

     as an estimate of (2). This is defined by $\mathrm{IMSE}(\xi) = \int_\chi E\left[\left\{\hat{Z}(\boldsymbol{x}) - \eta_1(\boldsymbol{x})\right\}^2\right]\mu(d\boldsymbol{x})$; in the Appendix we show that

$$\text{IMSE}\left(\xi\right)=\frac{\sigma^2}{n}tr\left(\boldsymbol{M}_\mu\boldsymbol{M}_\xi^{-1}\right)+\left\|\boldsymbol{M}_\mu^{1/2}\left(\boldsymbol{M}_\xi^{-1}\left(\boldsymbol{r}_\xi-\boldsymbol{s}_\xi\right)-\left(\boldsymbol{\theta}_0-\boldsymbol{\theta}_*\right)\right)\right\|^2+\tau^2. \tag{10}$$

We shall evaluate and compare $\text{IMSE}(\xi)$ and $\text{IMSE}\left(\tilde{\xi}\right)$.

- Briefly outline minimax?

1.1 Linear null models

- If the null model is linear, i.e. $\eta_0\left(\boldsymbol{x};\boldsymbol{\theta}\right)=\boldsymbol{f}'\left(\boldsymbol{x}\right)\boldsymbol{\theta}$, then $Y=Z$ and (1), (3) agree, and $\boldsymbol{f}\left(\boldsymbol{x}\right)$ does not depend on the parameters. Both (4) and (6) simplify: $\eta_0\left(\boldsymbol{x};\boldsymbol{\theta}_*\right)=\boldsymbol{f}'\left(\boldsymbol{x}\right)\boldsymbol{\theta}_*$, $\boldsymbol{M}_\mu^{-1}\boldsymbol{s}_\mu=\boldsymbol{\theta}_*$ and so, with $\boldsymbol{r}_\mu\stackrel{def}{=}\int_\chi\boldsymbol{f}\left(\boldsymbol{x}\right)\eta_1\left(\boldsymbol{x}\right)\mu\left(d\boldsymbol{x}\right)$,

$$\begin{aligned}\boldsymbol{\theta}_{0,\text{linear}}&=\boldsymbol{M}_\mu^{-1}\boldsymbol{r}_\mu,\\ \tau_{\text{linear}}^2&=\int_\chi\left(\eta_1\left(\boldsymbol{x}\right)-\boldsymbol{f}'\left(\boldsymbol{x}\right)\boldsymbol{\theta}_0\right)^2\mu\left(d\boldsymbol{x}\right)=\int_\chi\eta_1^2\left(\boldsymbol{x}\right)\mu\left(d\boldsymbol{x}\right)-\boldsymbol{\theta}_{0,\text{linear}}'\boldsymbol{M}_\mu\boldsymbol{\theta}_{0,\text{linear}}\end{aligned} \tag{11}$$

  Similarly, $\boldsymbol{M}_\xi^{-1}\boldsymbol{s}_\xi=\boldsymbol{M}_\mu^{-1}\boldsymbol{s}_\mu=\boldsymbol{\theta}_*$, and the BIAS and IMSE simplify:

$$\begin{aligned}\text{BIAS}\left[\hat{\boldsymbol{\theta}}\right]_{\text{linear}}&=\boldsymbol{M}_\xi^{-1}\boldsymbol{r}_\xi-\boldsymbol{M}_\mu^{-1}\boldsymbol{r}_\mu.\\ \text{IMSE}\left(\xi\right)_{\text{linear}}&=\frac{\sigma^2}{n}tr\left(\boldsymbol{M}_\mu\boldsymbol{M}_\xi^{-1}\right)+\left\|\boldsymbol{M}_\mu^{1/2}\left(\boldsymbol{M}_\xi^{-1}\boldsymbol{r}_\xi-\boldsymbol{\theta}_{0,\text{linear}}\right)\right\|^2+\tau^2.\end{aligned}$$

- Box and Draper (1959) consider linear null and alternative responses, i.e. $\eta_1\left(\boldsymbol{x}\right)=\boldsymbol{g}'\left(\boldsymbol{x}\right)\boldsymbol{\beta}$, with $\boldsymbol{g}\left(\boldsymbol{x}\right)=\left(\boldsymbol{f}'\left(\boldsymbol{x}\right),\boldsymbol{h}'\left(\boldsymbol{x}\right)\right)'$ so that (3) is nested within (2). They study designs under which $\text{BIAS}\left[\hat{\boldsymbol{\theta}}\right]_{\text{linear}}$ vanishes; a sufficient condition for this is $\int_\chi\boldsymbol{f}\left(\boldsymbol{x}\right)\boldsymbol{g}'\left(\boldsymbol{x}\right)\xi\left(d\boldsymbol{x}\right)=\int_\chi\boldsymbol{f}\left(\boldsymbol{x}\right)\boldsymbol{g}'\left(\boldsymbol{x}\right)\mu\left(d\boldsymbol{x}\right)$. For instance if the elements of $\boldsymbol{g}\left(x\right)$ are powers of x this requires the moments of the design to agree with those of the uniform distribution on χ up to a sufficiently high order. Box and Draper (1959) also investigate designs minimizing the IMSE; they are led to comment (p. 622) that "... the optimal design in typical situations in which both variance and bias occur is very nearly the same as would be obtained if *variance were ignored completely and the experiment designed so as to minimize the bias alone*".

2 Examples

2.1 Example 1: SLR, quadratic alternative

- This example somewhat revisits those of Box and Draper (1959) and Huber (1975). But in the former the emphasis was on constructing designs minimizing the IMSE at the alternate model, and in the latter on the construction of designs with giving minimax protection within a class of alternatives. Here we focus on the properties of the cluster designs.

- We take $\chi = [-1, 1]$, $\boldsymbol{f}(x) = (1, x)'$, $\eta_0(\boldsymbol{x}; \boldsymbol{\theta}) = \boldsymbol{f}'(x)\boldsymbol{\theta}$ and $\eta_1(x) = \frac{3\sqrt{5}}{2}\tau\left(x^2 - \frac{1}{3}\right)$, so that (11) holds with $\boldsymbol{r}_\mu = \boldsymbol{\theta}_0 = \boldsymbol{0}$ and

$$\boldsymbol{M}_\mu = \begin{pmatrix} 1 & 0 \\ 0 & \frac{1}{3} \end{pmatrix}.$$

We consider symmetric designs only, with $\xi = \{x_i, n_i/n\}_{i=1}^{I}$. Denote by

$$\nu_2 = \sum_{i=1}^{I} \frac{n_i}{n}\left(x_i^2 - \frac{1}{3}\right)$$

the difference between the second moment of the design, and that of μ. Then

$$\begin{aligned} \boldsymbol{M}_\xi &= \begin{pmatrix} 1 & 0 \\ 0 & \nu_2 \end{pmatrix}, \\ \boldsymbol{r}_\xi &= \frac{3\sqrt{5}}{2}\tau\begin{pmatrix} \nu_2 \\ 0 \end{pmatrix}, \end{aligned}$$

so that $\text{BIAS}\left[\hat{\boldsymbol{\theta}}\right]_{\text{linear}} = \boldsymbol{r}_\xi$ and

$$\begin{aligned} \text{IMSE}(\xi)_{\text{linear}} &= \frac{\sigma^2}{n} tr\left(\boldsymbol{M}_\mu \boldsymbol{M}_\xi^{-1}\right) + \left\|\boldsymbol{M}_\mu^{1/2}\boldsymbol{M}_\xi^{-1}\boldsymbol{r}_\xi\right\|^2 + \tau^2 \\ &= \frac{\sigma^2}{n}\left(1 + \frac{1}{3\nu_2}\right) + \tau^2\left(\frac{45}{4}\nu_2^2 + 1\right). \end{aligned}$$

- We compare ξ_0, which places half of the observations each endpoint of χ, with $\tilde{\xi}$, which spreads these replicates out into clusters of equally spaced points, over two subintervals $[-1, -1 + c]$ and $[1 - c, 1]$. The value of 'c' is chosen so that these two intervals together constitute $100p\%$ of χ, with p chosen by the user, For instance if $n = 10$ and $p = .1$, then $c = p$ and $\tilde{\xi} = \{\pm 0.900, \pm .925, \pm .950, \pm .975, \pm 1.00\}$. See Figure 1.

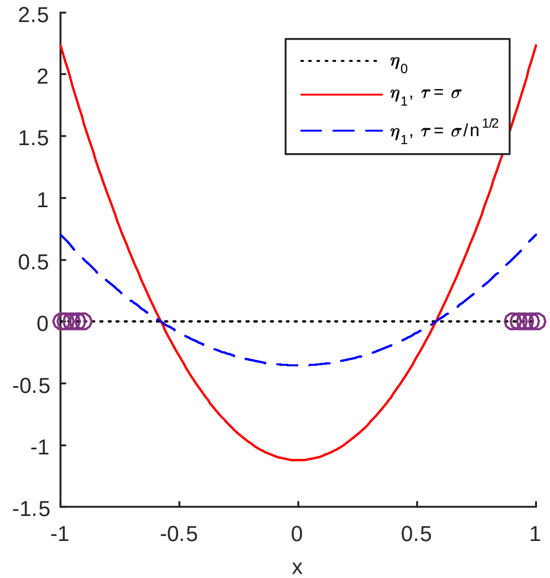

Figure 1: Response functions and design points of $\tilde{\xi}$.

- As well as $\text{IMSE}(\xi)$ we compute the losses $\phi_D = \left(\text{COV}\left[\hat{\boldsymbol{\theta}}\right]\right)^{1/2}$ and $\phi_I = tr\left(\boldsymbol{M}_\mu \text{COV}\left[\hat{\boldsymbol{\theta}}\right]\right)$, (integrated variance),both for ξ_0 and $\tilde{\xi}$. We then display the ratios $\text{IMSE}\left(\tilde{\xi}\right)/\text{IMSE}(\xi_0)$ and $\phi\left(\tilde{\xi}\right)/\phi\left(\xi_0\right)$ for $\phi \in \{\phi_D, \phi_I\}$.
- For instance, with $n = 10,\ p = .1, \tau = \sigma = 1$ we obtain

```
Losses =
                imse        phiD        phiI
     xi0     6.15000     0.12247     0.15000
xitilde      4.81891     0.13240     0.15844
  ratio      0.78356     1.08108     1.05625
```

With these parameters but $\tau = \sigma/\sqrt{n}$ only the first column changes:

```
                imse
     xi0     0.75000
xitilde      0.62448
  ratio      0.83265
```

- Would either of these designs detect a quadratic response of the form considered here? The F-test of the null hypothesis that the response is linear, vs. the alternate that it is quadratic, based on N observations has power

$$P\left(F^1_{n-3}\left(\lambda^2\right) > f_\alpha\right),$$

where $P\left(F^1_{n-3} > f_\alpha\right) = \alpha$ determines f_α and the non-centrality parameter (derived in the Appendix) is

$$\lambda^2\left(\xi\right) = \left(45/4\right) N \text{VAR}_\xi\left[X^2\right]\frac{\tau^2}{\sigma^2}. \tag{12}$$

Of course $\text{VAR}_{\xi_0}\left[X^2\right] = 0$ and there is no power. With $N = n = 10$ as above, $\text{VAR}_{\tilde{\xi}}\left[X^2\right] = .0045$ and the powers of the test, with size $\alpha = .05$, and $\tau = \sigma$ or $\sigma/\sqrt{n}$ are only .095 and .054, respectively. These improve if p is increased.

- To maximize the power of this test a design should maximize $\text{VAR}_\xi\left[X^2\right]$, this is the design with masses $(.25, .50, .25)$ at $(-1, 0, -1)$. Putting this amount of mass at 0 is somewhat extreme; a reasonable 'compromise' design, for efficiency of estimation and a reasonable power in this test, would be the D-optimal design for quadratic regression, placing mass 1/3 at each of $(-1, 0, 1)$. In fact this is the solution of Studden (1982) to the problem of maximizing the power of the test, subject to the requirement that the variance of the estimate of the quadratic coefficient be no more than 9/8 that of the optimal design for estimating this coefficient. We consider this D-optimal design next.

2.2 Example 2: Quadratic null, nonlinear alternative

2.3 Example 3: MM null, exponential alternative

- Source: Biedermann and Yang (2015)

- Null (i.e., experimenter's fitted) model is Michaelis-Menten:

$$E\left[Y|x\right] = \eta_0\left(\boldsymbol{x};\boldsymbol{\theta}\right) = \frac{\theta_1 x}{\theta_2 + x}, \theta_1, \theta_2 > 0, 0 \le x \le B;$$

here we have set $\boldsymbol{\theta} = (\theta_1, \theta_2)'$. The alternate model will be either this MM model with an incorrect initial value, or the exponential response

$$\eta_1\left(x;\boldsymbol{\phi}\right) = \phi_1\left(1 - e^{-\phi_2 x}\right),$$

where $\boldsymbol{\phi} = (\phi_1, \phi_2)'$, or both. We tentatively take $B = 400$ and $\boldsymbol{\phi} = (.1, 1)'$.

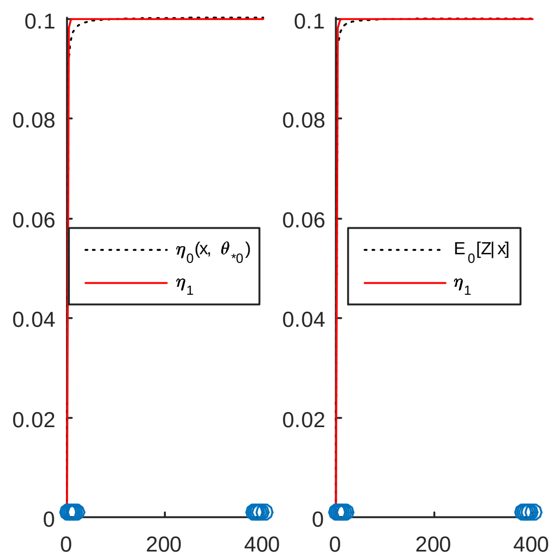

Figure 2: Michaelis-Menten and exponential responses. Left: optimal local value $\boldsymbol{\theta}_{*,0}$; right: local value $\boldsymbol{\theta}_0 \neq \boldsymbol{\theta}_{*,0}$.

- For a design ξ with I points of support, placing n_i observations at x_i $(\sum_{i=1}^{I} n_i = n)$, the information matrix is

$$\boldsymbol{M}\left(\xi; \boldsymbol{\theta}_*\right) = \frac{1}{n}\sum_{i=1}^{I} n_i \boldsymbol{f}\left(x_i\right)\boldsymbol{f}'\left(x_i\right),$$

where

$$\boldsymbol{f}\left(x\right) = \dot{\eta}_0\left(x; \boldsymbol{\theta}_*\right) = \left(\begin{array}{c} \frac{x}{\theta_2+x} \\ -\frac{\theta_1 x}{(\theta_2+x)^2} \end{array}\right)_{|\boldsymbol{\theta}_*} = \frac{x}{\theta_2+x}\left(\begin{array}{c} 1 \\ -\frac{\theta_1}{\theta_2+x} \end{array}\right)_{|\boldsymbol{\theta}_*}.$$

- For a design of size $n = 2m$, the D-optimal (local) design ξ_0 has

$$\begin{aligned} x_1 &= \cdots . x_m = \frac{B\theta_{*,2}}{B+2\theta_{*,2}}, \\ x_{m+1} &= \cdots = x_n = B. \end{aligned}$$

- We first compute

$$\begin{aligned} \boldsymbol{\theta}_{*,0} &= \arg\min_{\boldsymbol{\theta}_*} \int_{\chi} \left\{\eta_1\left(\boldsymbol{x}\right) - \eta_0\left(\boldsymbol{x}; \boldsymbol{\theta}_*\right)\right\}^2 \mu\left(d\boldsymbol{x}\right), \\ \tau_0^2 &= \int_{\chi} \left\{\eta_1\left(\boldsymbol{x}\right) - \eta_0\left(\boldsymbol{x}; \boldsymbol{\theta}_{*,0}\right)\right\}^2 \mu\left(d\boldsymbol{x}\right), \end{aligned}$$

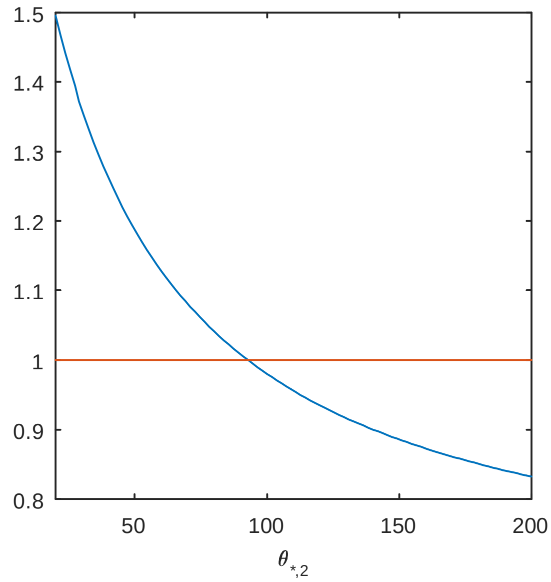

Figure 3: Ratios $\phi_D\left(\tilde{\xi}_1|\theta_{*,2}\right)/\phi_D\left(\xi_1|\theta_{*,2}\right)$.

the efficiency measures based on $\boldsymbol{M}(\xi)$ and

$$\text{IMSE}(\xi) = \frac{\sigma^2}{n} tr\left(\boldsymbol{M}_\mu \boldsymbol{M}_\xi^{-1}\right) + \left\|\boldsymbol{M}_\mu^{1/2}\boldsymbol{M}_\xi^{-1}\left(\boldsymbol{r}_\xi - \boldsymbol{s}_\xi\right)\right\|^2 + \tau_0^2,$$

at ξ_0 and $\tilde{\xi}$, using the 'perfect' starting value $\boldsymbol{\theta}_* = \boldsymbol{\theta}_{*,0}$. This tests the robustness of the designs against model misspecification alone. In this case, using the values given above, $\boldsymbol{\theta}_* = \boldsymbol{\theta}_{*,0} = (.1004, .3919)'$, we have $\tau_0 = .0231$, the design ξ_0 places mass of .5 at each of .390 and 400, and using $p = .1$, $\tilde{\xi}$ has design points .39(5)20.39 and 380(5)400. (**Note**: all are *inside* of the design points of ξ_0; this seems to make a big difference.) The output is

	imse	phiD	phiI
xi0	0.19618	3.13075	0.19564
xitilde0	0.12519	5.28728	0.12465
ratio	0.63814	1.68882	0.63714

- We then take an 'incorrect' starting value $\boldsymbol{\theta}_* = (.1, .05)' \neq \boldsymbol{\theta}_{*,0}$, for which $\boldsymbol{\theta}_0 = (.1002, .2364)'$ from (4) and $\tau = .0252$ from (6). We then again compute the efficiency measures based on $\boldsymbol{M}(\xi)$ and IMSE(ξ) at (10) at ξ_0, with mass of .5 at each of .05 and 400 and the corresponding clustered design $\tilde{\xi}$ with design points .05(5)20.05 and 380(5)400. This tests the robustness of the designs against both model misspecification and incorrect starting values. See Figure 2. The output is

	imse	phiD	phiI
xi0	0.19978	0.40015	0.19914
xitilde0	0.11336	0.67089	0.11273
ratio	0.56745	1.67659	0.56606

- The robustness is attained both against the model form and the choice of starting value. To see the effect of the starting value alone we took $\eta_1(x) = \eta_0(x; \boldsymbol{\phi})$, an incorrect starting value $\boldsymbol{\theta}_* = (.1, .05)' \neq \boldsymbol{\phi}$, computed $\boldsymbol{\theta}_0 = (.1, .3049)'$ from (4) and $\tau = .0485$. We compare the D-optimal design ξ_0 with the corresponding cluster design $\tilde{\xi}$, both of which are the same as in the previous bullet, obtaining a change in the first column of the output:

	imse
xi0	0.20149
xitilde0	0.11508
ratio	0.57115

- We also computed the 'standardized maximin' design ξ_1 which is 'optimally robust' in the classical sense, against misspecified starting values. This has been derived in Dette and Biedermann (2003), and is described as follows. Suppose that one seeks robustness against a range of initial values $\theta_{*,2} \in \left[\theta_2^{(0)}, \theta_2^{(1)}\right]$. (Since the model is conditionally linear in θ_1, initial values for this parameter are not needed.) Put $\beta_0 = \theta_2^{(0)}/B$, $\beta_1 = \theta_2^{(1)}/B$. Then ξ_1 has equal masses at B and

$$\begin{aligned} x_1 &= B \cdot \frac{\beta_1\sqrt{\beta_0(1+\beta_0)} - \beta_0\sqrt{\beta_1(1+\beta_1)}}{\sqrt{\beta_1(1+\beta_1)} - \sqrt{\beta_0(1+\beta_0)}} \\ &= \frac{\theta_2^{(1)}\sqrt{\theta_2^{(0)}\left(B+\theta_2^{(0)}\right)} - \theta_2^{(0)}\sqrt{\theta_2^{(1)}\left(B+\theta_2^{(1)}\right)}}{\sqrt{\theta_2^{(1)}\left(B+\theta_2^{(1)}\right)} - \sqrt{\theta_2^{(0)}\left(B+\theta_2^{(0)}\right)}}. \end{aligned}$$

For instance if $\theta_2^{(0)} = \theta_2^{(1)} = \theta_{*,2}$, l'Hospital's Rule gives $x_1 = B\theta_{*,2}/B + 2\theta_{*,2}$, as must be the case.

Using $B = 400$ as above, we computed the maximin design for the range $\left[\theta_2^{(0)}, \theta_2^{(1)}\right] = [.05B, .50B]$. Apart from the value of B this is as for case 'B' of Example 1 in Dette and Biedermann (2003). We obtained $x_1 = 44.76$. We compared ξ_1 with the design $\tilde{\xi}_1$, computed as in the previous bullet but with clustered based on the D-optimal design with initial value $\boldsymbol{\theta}_* = (.1, \left(\theta_2^{(0)} + \theta_2^{(1)}\right)/2)' = (.1, 225)$. We found that the support points of $\tilde{\xi}_1$

are 105.88(5)125.88 and 380(5)400. If these designs are then employed with initial values $\theta_{*,2} \in \left[\theta_2^{(0)}, \theta_2^{(1)}\right]$, then the ratios of maximum losses (recall that 'loss' is the root of the determinant of the covariance matrix), i.e. $\phi_D\left(\tilde{\xi}_1|\theta_{*,2}\right)/\phi_D\left(\xi_1|\theta_{*,2}\right)$, are pictured in Figure 3, and are < 1 over most of the range of $\theta_{*,2}$.

Appendix: Derivations

Derivation of (10). Note that a consequence of the minimization defining $\boldsymbol{\theta}_0$ is that

$$\int_{\chi} \boldsymbol{f}\left(\boldsymbol{x}\right)\left\{\eta_1\left(\boldsymbol{x}\right) - \eta_0\left(\boldsymbol{x};\boldsymbol{\theta}_*\right) - \boldsymbol{f}'\left(\boldsymbol{x}\right)\left(\boldsymbol{\theta}_0 - \boldsymbol{\theta}_*\right)\right\}^2 \mu\left(d\boldsymbol{x}\right) = \boldsymbol{0}. \tag{A.1}$$

This allows IMSE to be partitioned into, first two and then three, orthogonal components as

$$\begin{aligned}
\text{IMSE}\left(\xi\right) &= \int_{\chi} E\left[\left\{\eta_1\left(\boldsymbol{x}_1\right) - \eta_0\left(\boldsymbol{x};\boldsymbol{\theta}_*\right) - \boldsymbol{f}'\left(\boldsymbol{x}\right)\left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_*\right)\right\}^2\right]\mu\left(d\boldsymbol{x}\right) \\
&= \int_{\chi} E\left[\left\{\eta_1\left(\boldsymbol{x}\right) - \eta_0\left(\boldsymbol{x};\boldsymbol{\theta}_*\right) - \boldsymbol{f}'\left(\boldsymbol{x}\right)\left(\boldsymbol{\theta}_0 - \boldsymbol{\theta}_*\right) - \boldsymbol{f}'\left(\boldsymbol{x}\right)\left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right)\right\}^2\right]\mu\left(d\boldsymbol{x}\right) \\
&= \int_{\chi} E\left[\left\{\eta_1\left(\boldsymbol{x}\right) - \eta_0\left(\boldsymbol{x};\boldsymbol{\theta}_*\right) - \boldsymbol{f}'\left(\boldsymbol{x}\right)\left(\boldsymbol{\theta}_0 - \boldsymbol{\theta}_*\right)\right\}^2\right]\mu\left(d\boldsymbol{x}\right) \\
&\quad + \int_{\chi} E\left[\boldsymbol{f}'\left(\boldsymbol{x}\right)\left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right)^2\right]\mu\left(d\boldsymbol{x}\right) \\
&= \tau^2 + tr\left(\boldsymbol{M}_{\mu}\left\{\text{COV}\left[\hat{\boldsymbol{\theta}}\right] + \text{BIAS}\left[\hat{\boldsymbol{\theta}}\right]\text{BIAS}\left[\hat{\boldsymbol{\theta}}\right]'\right\}\right).
\end{aligned}$$

Substituting (8) and (9) gives (10). □

Derivation of (12). We consider symmetric designs, on the basis of which N observations are made (so even clustered designs $\tilde{\xi}$ might be replicated). Let $x_1, ..., x_N$ be the, not necessarily unique, design points. The experimenter fits a quadratic model, so that the mean vector of the data $\boldsymbol{Y}_{N\times 1}$ is

$$\boldsymbol{\mu}_{N\times 1} = \boldsymbol{X}\boldsymbol{\theta} = \left(\boldsymbol{1}\vdots\boldsymbol{u}\vdots\boldsymbol{v}\right)\boldsymbol{\theta} \stackrel{def}{=} \left(\boldsymbol{X}_1\vdots\boldsymbol{v}\right)\begin{pmatrix}\boldsymbol{\theta}_1 \\ \theta_2\end{pmatrix},$$

where $\boldsymbol{u} = (x_1, ..., x_N)'$, $\boldsymbol{v} = (x_1^2, ..., x_N^2)$. The noncentrality parameter of the F-statistic used to test the hypothesis that $\theta_2 = 0$ is

$$\lambda^2 = \frac{\min_{\boldsymbol{t}}\left\|\boldsymbol{X}\boldsymbol{\theta} - \boldsymbol{X}_1\boldsymbol{t}\right\|^2}{\sigma^2}.$$

The minimizing $\boldsymbol{t}$ satisfies $\boldsymbol{X}_1' \left(\boldsymbol{X}\boldsymbol{\theta} - \boldsymbol{X}_1\boldsymbol{t} \right) = \boldsymbol{0}$, whence $\boldsymbol{t} = \left(\boldsymbol{X}_1'\boldsymbol{X}_1 \right)^{-1} \boldsymbol{X}_1'\boldsymbol{X}\boldsymbol{\theta}$ and $\sigma^2\lambda^2 = \left\| \left(\boldsymbol{I} - \boldsymbol{H}_1 \right) \boldsymbol{X}\boldsymbol{\theta} \right\|^2$, where $\boldsymbol{H}_1 = \boldsymbol{X}_1 \left(\boldsymbol{X}_1'\boldsymbol{X}_1 \right)^{-1} \boldsymbol{X}_1'$. Put $S_2 = \sum_{i=1}^N x_i^2$, $S_4 = \sum_{i=1}^N x_i^4$. From

$$\begin{aligned}
\left(\boldsymbol{I} - \boldsymbol{H}_1 \right) \boldsymbol{X}\boldsymbol{\theta} &= \left(\boldsymbol{I} - \boldsymbol{H}_1 \right) \left(\boldsymbol{X}_1 \vdots \boldsymbol{v} \right) \boldsymbol{\theta} = \left(\boldsymbol{0} \vdots \left(\boldsymbol{I} - \boldsymbol{H}_1 \right) \boldsymbol{v} \right) \boldsymbol{\theta} = \theta_2 \left(\boldsymbol{I} - \boldsymbol{H}_1 \right) \boldsymbol{v}, \\
\text{and } \boldsymbol{v}'\boldsymbol{H}_1\boldsymbol{v} &= \boldsymbol{v}'\boldsymbol{X}_1 \left(\boldsymbol{X}_1'\boldsymbol{X}_1 \right)^{-1} \boldsymbol{X}_1'\boldsymbol{v} = (S_2, 0) \begin{pmatrix} N & 0 \\ 0 & S_2 \end{pmatrix}^{-1} \begin{pmatrix} S_2 \\ 0 \end{pmatrix} = \frac{S_2^2}{N},
\end{aligned}$$

we obtain $\sigma^2\lambda^2 = \theta_2^2 \boldsymbol{v}' \left(\boldsymbol{I} - \boldsymbol{H}_1 \right) \boldsymbol{v} = \theta_2^2 \left\{ S_4 - \frac{S_2^2}{N} \right\} = N\theta_2^2 \mathrm{VAR}_\xi \left[X^2 \right]$. Since $\eta_1(x)$ has $\theta_2 = \frac{3\sqrt{5}}{2}\tau$, the power of the test is based on (12). □

Acknowledgements

This work was carried out with the support of the Natural Sciences and Engineering Council of Canada.

References

Biedermann, S. and Yang, M. (2015); Designs for Selected Noninear Models, Chapter14 of *Handbook of Design and Analysis of Experiments*, Chapman & Hall/CRC.

Box, G.E.P., and Draper, N.R. (1959), "A Basis for the Selection of a Response Surface Design," *Journal of the American Statistical Association*, 54, 622-654.

Dette, H., and Biedermann, S. (2003), "Robust and Efficient Designs for the Michaelis–Menten Model," *Journal of the American Statistical Association*, 98, 679-686.

Huber, P. J. (1975), "Robustness and Designs," in: *A Survey of Statistical Design and Linear Models*, ed. J. N. Srivastava, Amsterdam: North Holland, pp. 287-303.

Studden, W. J. (1982), "Some Robust-Type *D*-Optimal Designs in Polynomial Regression," *Journal of the American Statistical Association*, 77, 916-921.