Minimax Robust Designs for M-Estimated Models

Rui Hu[a,*], Douglas P. Wiens[b,**]

[a]*Mathematics & Statistics, MacEwan University, Edmonton, Canada, T5J 4S2*
[b]*Mathematical & Statistical Sciences, University of Alberta, Edmonton, Canada, T6G 2G1*

This draft March 11, 2026

Abstract

Designs which are minimax in the presence of model misspecifications have been constructed so as to minimize the maximum, over classes of alternate response models, of the integrated mean squared error of the predicted values. The theory to date has focussed almost exclusively on Least Squares estimates. Here we extend this theory to designs tailored for M-estimation of parameters, thus obtaining additional robustness against outlying responses. We show that, subject to a minor change in a tuning constant, designs optimal for Least Squares remain so asymptotically for M-estimation. We argue that even this minor change should be ignored, and the tuning constant chosen in an *ad hoc* but sensible manner which does not depend on which M-estimate is being employed.

Keywords: asymptotics, finite design space, misspecified model, regression design.
2010 MSC: Primary 62F35, Secondary 62K05

1. Introduction and summary

The theory and practice of robustness of design, for possibly misspecified response functions, is well-developed as it applies to cases in which parameter estimation is to be carried out by Least Squares (LS). An investigator seeking model robustness might naturally be concerned as well with robustness against outlying data points, or more generally against a misspecified data-generating probability distribution, and hence seek M-estimates of the parameters. There is little guidance furnished in the literature as regards appropriate designs in this case. Wiens (1994, 1996) studied this design problem for quite limited classes of approximate responses on continuous design spaces and obtained asymptotic results under rather restrictive conditions. Wiens and Wu (2010) carried out a small simulation study and found that there was little dependence of the designs on the method of estimation.

In this article we strengthen these results. We show that, if the parameters of the assumed model are to be estimated by Ordinary M-estimation, and if – as is very common

*E-mail: hur3@macewan.ca.
**E-mail: doug.wiens@ualberta.ca. Supplementary material is at www.ualberta.ca/~dwiens/.

– the design space is finite, then designs optimal for LS remain so, asymptotically, for M-estimation.

We note that M-estimates offer protection against outlying responses, but not necessarily against outliers in the factor space. This latter type of protection is furnished by Generalized M-estimation. But since our focus centres on design points chosen by the experimenter, such outlyingness is not an issue.

In §2.1 of this article we give the asymptotic theory on which our design problem will be based. Then in §2.2 we address the design problem. We show there that the optimally robust designs depend on the anticipated method of estimation only through a tuning constant, which one might interpret as the relative emphasis placed by the designer on bias reduction versus variance reduction. Examples and methods of implementation are studied and discussed in §3. We argue there that, although the aforementioned tuning constant depends on the method of estimation through unknown parameters, this dependence is so slight that the previously mentioned, more *ad hoc* method of choosing this constant should be used. The result is then that, given this constant, the designs are completely independent of which M-estimate is to be used.

Proofs are in the Appendix. The MATLAB code used to prepare the examples is available on the second author's personal website.

2. Minimax robustness of design

2.1. Asymptotic theory

Our minimax design problem is phrased in terms of an approximate regression response

$$E\left[Y(\boldsymbol{x})\right] \approx \boldsymbol{f}'(\boldsymbol{x})\,\boldsymbol{\theta}, \tag{1}$$

for p regressors $\boldsymbol{f}$, each functions of q independent variables $\boldsymbol{x}$ ranging over a *design space* $\mathcal{X} \subset \mathbb{R}^q$, and a parameter $\boldsymbol{\theta}$. At such values of $\boldsymbol{x}$, $Y(\boldsymbol{x})$ is observed with additive random error: $Y(\boldsymbol{x}) = E\left[Y(\boldsymbol{x})\right] + \varepsilon(\boldsymbol{x})$, for i.i.d. errors ε (implying that the distribution of $\varepsilon(\boldsymbol{x})$ does not depend on $\boldsymbol{x}$). The error distribution is assumed to be symmetric about 0.

Given observations $\left\{Y_j(\boldsymbol{x}_i)\,|\,j = 1, ..., n_i\right\}$ at distinct points $\boldsymbol{x}_i$, we suppose that $\boldsymbol{\theta}$ will be estimated by M-estimation with an auxiliary estimate of scale: for a convex, even, non-negative function ρ, with absolutely continuous derivative ψ:

$$\begin{aligned}
\hat{\boldsymbol{\theta}}_n &= \arg\min_{\boldsymbol{\theta}} \sum_{i,j} \rho\left(\frac{Y_j(\boldsymbol{x}_i) - \boldsymbol{f}'(\boldsymbol{x}_i)\,\boldsymbol{\theta}}{\hat{\sigma}_n}\right), \\
\hat{\sigma}_n &= \mathit{median}\left\{\left|Y_j(\boldsymbol{x}_i) - \boldsymbol{f}'(\boldsymbol{x}_i)\,\hat{\boldsymbol{\theta}}_n\right|\right\} / \Phi^{-1}(.75)\,.
\end{aligned}$$

Let σ be the limit in probability of $\hat{\sigma}_n$. The factor $\Phi^{-1}(.75)$ ensures that σ is the standard deviation of the random errors if these are normally distributed. Under the mild conditions of Rocke and Shannon (1986), $\hat{\boldsymbol{\theta}}_n$ has the same asymptotic properties as if $\hat{\sigma}_n$ were replaced by σ.

Since (1) is an approximation the interpretation of $\boldsymbol{\theta}$ is unclear; we *define* this target parameter by

$$\boldsymbol{\theta} = \arg\min_{\boldsymbol{\eta}} \int_{\mathcal{X}} \left(E\left[Y\left(\boldsymbol{x}\right)\right] - \boldsymbol{f}'\left(\boldsymbol{x}\right)\boldsymbol{\eta}\right)^2 \mu\left(d\boldsymbol{x}\right), \tag{2}$$

where $\mu\left(d\boldsymbol{x}\right)$ represents either Lebesgue measure or counting measure, depending upon the nature of the design space $\mathcal{X}$ with $\int_{\mathcal{X}} \mu\left(d\boldsymbol{x}\right) < \infty$. Equivalently, and with $\tau\left(\boldsymbol{x}\right) \stackrel{def}{=} E\left[Y\left(\boldsymbol{x}\right)\right] - \boldsymbol{f}'\left(\boldsymbol{x}\right)\boldsymbol{\theta}$, we have

$$\int_{\mathcal{X}} \boldsymbol{f}\left(\boldsymbol{x}\right)\tau\left(\boldsymbol{x}\right)\mu\left(d\boldsymbol{x}\right) = \boldsymbol{0}. \tag{3}$$

Assuming that $\mathcal{X}$ is rich enough that the matrix $\boldsymbol{A} = \int_{\mathcal{X}} \boldsymbol{f}\left(\boldsymbol{x}\right)\boldsymbol{f}'\left(\boldsymbol{x}\right)\mu\left(d\boldsymbol{x}\right)$ is invertible, the parameter defined by (2) and (3) is unique. We bound the approximation error in (1) by assuming that

$$\int_{\mathcal{X}} \tau^2\left(\boldsymbol{x}\right)\mu\left(d\boldsymbol{x}\right) \le \kappa^2/n, \tag{4}$$

for a constant κ.

Our model is now given by $E\left[Y\left(\boldsymbol{x}\right)\right] = \boldsymbol{f}'\left(\boldsymbol{x}\right)\boldsymbol{\theta} + \tau\left(\boldsymbol{x}\right)$, for an unknown model error $\tau\left(\cdot\right)$ constrained by (3) and (4).

We identify a design with its design measure – a probability measure $\xi_n\left(d\boldsymbol{x}\right)$ on $\mathcal{X}$. Define

$$\boldsymbol{M}_0\left(\xi\right) = \int_{\mathcal{X}} \boldsymbol{f}\left(\boldsymbol{x}\right)\boldsymbol{f}'\left(\boldsymbol{x}\right)\xi\left(d\boldsymbol{x}\right), \qquad \boldsymbol{b}_0\left(\xi\right) = \int_{\mathcal{X}} \boldsymbol{f}\left(\boldsymbol{x}\right)\tau\left(\boldsymbol{x}\right)\xi\left(d\boldsymbol{x}\right),$$

and set $\boldsymbol{M}_{0,n} = \boldsymbol{M}_0\left(\xi_n\right)$, $\boldsymbol{b}_{0,n} = \boldsymbol{b}_0\left(\xi_n\right)$. Assume ξ_n is such that $\boldsymbol{M}_{0,n}$ is invertible. Finally, define

$$\sigma_M^2 = \sigma^2 E\left[\psi^2\left(\frac{\varepsilon}{\sigma}\right)\right] \Big/ \left(E\left[\psi'\left(\frac{\varepsilon}{\sigma}\right)\right]\right)^2. \tag{5}$$

In the appendix we prove Theorem 1, which asserts that $\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}$ is asymptotically normal:

$$\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta} \sim AN\left(\boldsymbol{M}_{0,n}^{-1}\boldsymbol{b}_{0,n}, \frac{\sigma_M^2}{n}\boldsymbol{M}_{0,n}^{-1}\right). \tag{6}$$

Theorem 1. *With notation as above, we have that, as $n \to \infty$,*

$$\sqrt{n}\boldsymbol{M}_{0,n}^{1/2}\left(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta} - \boldsymbol{M}_{0,n}^{-1}\boldsymbol{b}_{0,n}\right) \xrightarrow{d} N\left(\boldsymbol{0}, \sigma_M^2 \boldsymbol{I}_p\right).$$

2.2. Minimax theory

We now assume that $\mu\left(d\boldsymbol{x}\right)$ is counting measure on a finite design space $\chi = \{\boldsymbol{x}_1, ..., \boldsymbol{x}_N\}$. If a design ξ allocates n_i runs to $\boldsymbol{x}_i$, then $\xi_i \stackrel{def}{=} \xi\left(\boldsymbol{x}_i\right) = n_i/n$.

Let Υ be the class of functions $\tau\left(\cdot\right)$ defined by (3) and (4), which now take the form

$$\text{(a)} \sum_{i=1}^{N} \boldsymbol{f}\left(\boldsymbol{x}_i\right)\tau\left(\boldsymbol{x}_i\right) = \boldsymbol{0}, \quad \text{(b)} \sum_{i=1}^{N} \tau^2\left(\boldsymbol{x}_i\right) \le \kappa^2/n. \tag{7}$$

With $\hat{Y}(\boldsymbol{x}) = \boldsymbol{f}'(\boldsymbol{x})\hat{\boldsymbol{\theta}}_n$ we define our loss in terms of Integrated Mean Squared Error, maximized over $\tau(\cdot)$:

$$\text{IMSE}(\xi|\tau) = \sum_{i=1}^{N} E\left\{\left(E[Y(\boldsymbol{x}_i)] - \hat{Y}(\boldsymbol{x}_i)\right)^2\right\} = \sum_{i=1}^{N} E\left\{\left(\tau(\boldsymbol{x}_i) - \boldsymbol{f}'(\boldsymbol{x}_i)\left(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}\right)\right)^2\right\}. \quad (8)$$

Expanding (8) and using (3) and (6) results in

$$\begin{aligned}
\text{IMSE} &= \sum_{i=1}^{N} \tau^2(\boldsymbol{x}_i) + \sum_{i=1}^{N} \boldsymbol{f}'(\boldsymbol{x}_i) E\left[\left(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}\right)\left(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}\right)'\right] \boldsymbol{f}(\boldsymbol{x}_i) \\
&= \sum_{i=1}^{N} \tau^2(\boldsymbol{x}_i) + \sum_{i=1}^{N} \boldsymbol{f}'(\boldsymbol{x}_i)\left[\boldsymbol{M}_{0,n}^{-1}\boldsymbol{b}_{0,n}\boldsymbol{b}_{0,n}'\boldsymbol{M}_{0,n}^{-1} + \frac{\sigma_M^2}{n}\boldsymbol{M}_{0,n}^{-1}\right]\boldsymbol{f}(\boldsymbol{x}_i) + o\left(n^{-1}\right) \\
&= \sum_{i=1}^{N} \tau^2(\boldsymbol{x}_i) + \boldsymbol{b}_{0,n}'\boldsymbol{M}_{0,n}^{-1}\mathbf{A}\boldsymbol{M}_{0,n}^{-1}\boldsymbol{b}_{0,n} + \frac{\sigma_M^2}{n} tr\boldsymbol{A}\boldsymbol{M}_{0,n}^{-1} + o\left(n^{-1}\right),
\end{aligned}$$

with $\boldsymbol{A} = \sum_{i=1}^{N} \boldsymbol{f}(\boldsymbol{x}_i)\boldsymbol{f}'(\boldsymbol{x}_i)$.

To maximize IMSE over τ, note that both of the first two terms become larger if $\tau(\boldsymbol{x})$ is multiplied by a constant exceeding one, hence at a maximum (7b) is attained with equality. Define

$$\tau_0(\boldsymbol{x}) = \sqrt{n}\tau(\boldsymbol{x})/\kappa \text{ and } \boldsymbol{c}(\xi) = \sum_{i=1}^{N} \boldsymbol{f}(\boldsymbol{x}_i)\tau_0(\boldsymbol{x}_i)\xi(\boldsymbol{x}_i) = \sqrt{n}\boldsymbol{b}(\xi)/\kappa.$$

Then with $\boldsymbol{c}_n = \boldsymbol{c}(\xi_n)$ and

$$\nu = \kappa^2/\left(\sigma_M^2 + \kappa^2\right), \quad (9)$$

we have that $\max_{\tau\in\Upsilon}\text{IMSE}(\xi_n|\tau)$ is $\left(\sigma_M^2 + \kappa^2\right)/n$ times

$$I_\nu(\xi_n) = (1-\nu)\, tr\boldsymbol{A}\boldsymbol{M}_{0,n}^{-1} + \nu\left(1 + \max_{\tau_0} \boldsymbol{c}_n'\boldsymbol{M}_{0,n}^{-1}\mathbf{A}\boldsymbol{M}_{0,n}^{-1}\boldsymbol{c}_n\right), \quad (10)$$

with τ_0 constrained by (7a) and by $\sum_{i=1}^{N} \tau_0^2(\boldsymbol{x}_i) = 1$.

We seek a minimax design ξ_n^*, defined as

$$\xi_n^* = \arg\min_{\xi_n} I_\nu(\xi_n).$$

To carry out the maximization in (10) it is convenient to introduce an orthogonal basis for the space of regressors. Define $\boldsymbol{F}_{N\times p} = (\boldsymbol{f}(\boldsymbol{x}_1), \cdots, \boldsymbol{f}(\boldsymbol{x}_N))'$, $\boldsymbol{\tau}_0 = (\tau_0(\boldsymbol{x}_1), \cdots, \tau_0(\boldsymbol{x}_N))'$, and for a design ξ on χ, $\boldsymbol{D}(\xi) = diag(\xi_1, \cdots, \xi_N)$. By the Gram-Schmidt process we can construct a matrix $\boldsymbol{Q}_{N\times p}$ whose orthonormal columns form a basis for the column space of $\boldsymbol{F}$ – assumed to be of dimension p – and then $\boldsymbol{F} = \boldsymbol{QT}$ for a non-singular lower triangular matrix $\boldsymbol{T}$. Let $\boldsymbol{Q}_\perp : N \times N - p$ be the orthogonal complement of $\boldsymbol{Q}$, so that $\left(\boldsymbol{Q} \vdots \boldsymbol{Q}_\perp\right) : N \times N$

is orthogonal. Condition (7a) asserts that $\boldsymbol{\tau}_0$ is orthogonal to the columns of $\boldsymbol{Q}$, hence is a linear combination of the columns of $\boldsymbol{Q}_\perp$ and so there is $\boldsymbol{\beta}_{N-p\times 1}$ of unit norm for which $\boldsymbol{\tau}_0 = \boldsymbol{Q}_\perp \boldsymbol{\beta}$.

In this notation $\boldsymbol{A} = \boldsymbol{F}'\boldsymbol{F} = \boldsymbol{T}'\boldsymbol{T}$ and

$$\begin{aligned} \boldsymbol{M}_{0,n} &= \boldsymbol{F}'\boldsymbol{D}(\xi_n)\boldsymbol{F} = \boldsymbol{T}'\boldsymbol{Q}'\boldsymbol{D}(\xi_n)\boldsymbol{Q}\boldsymbol{T}, \\ \boldsymbol{c}_n &= \boldsymbol{F}'\boldsymbol{D}(\xi_n)\boldsymbol{\tau}_0 = \boldsymbol{T}'\boldsymbol{Q}'\boldsymbol{D}(\xi_n)\boldsymbol{Q}_{\perp\boldsymbol{\beta}}; \end{aligned}$$

then (10) becomes

$$I_\nu(\xi_n) = (1-\nu)\, tr\left(\boldsymbol{Q}'\boldsymbol{D}(\xi_n)\boldsymbol{Q}\right)^{-1} + \nu\left(1 + \max_{\|\boldsymbol{\beta}\|=1} \boldsymbol{\beta}'\boldsymbol{Q}_\perp'\boldsymbol{D}(\xi_n)\boldsymbol{Q}\left(\boldsymbol{Q}'\boldsymbol{D}(\xi_n)\boldsymbol{Q}\right)^{-2}\boldsymbol{Q}'\boldsymbol{D}(\xi_n)\boldsymbol{Q}_\perp\boldsymbol{\beta}\right). \tag{11}$$

With $ch_{\max}$ denoting the maximum eigenvalue we have

$$\begin{aligned} & \max_{\|\boldsymbol{\beta}\|=1} \boldsymbol{\beta}'\boldsymbol{Q}_\perp'\boldsymbol{D}(\xi_n)\boldsymbol{Q}\left(\boldsymbol{Q}'\boldsymbol{D}(\xi_n)\boldsymbol{Q}\right)^{-2}\boldsymbol{Q}'\boldsymbol{D}(\xi_n)\boldsymbol{Q}_\perp\boldsymbol{\beta} \\ = \; & ch_{\max}\boldsymbol{Q}_\perp'\boldsymbol{D}(\xi_n)\boldsymbol{Q}\left(\boldsymbol{Q}'\boldsymbol{D}(\xi_n)\boldsymbol{Q}\right)^{-1}\cdot\left(\boldsymbol{Q}'\boldsymbol{D}(\xi_n)\boldsymbol{Q}\right)^{-1}\boldsymbol{Q}'\boldsymbol{D}(\xi_n)\boldsymbol{Q}_\perp \\ = \; & ch_{\max}\left(\boldsymbol{Q}'\boldsymbol{D}(\xi_n)\boldsymbol{Q}\right)^{-1}\boldsymbol{Q}'\boldsymbol{D}(\xi_n)\boldsymbol{Q}_\perp\boldsymbol{Q}_\perp'\boldsymbol{D}(\xi_n)\boldsymbol{Q}\left(\boldsymbol{Q}'\boldsymbol{D}(\xi_n)\boldsymbol{Q}\right)^{-1} \\ = \; & ch_{\max}\left(\boldsymbol{Q}'\boldsymbol{D}(\xi_n)\boldsymbol{Q}\right)^{-1}\boldsymbol{Q}'\boldsymbol{D}(\xi_n)\left(\boldsymbol{I}_N - \boldsymbol{Q}\boldsymbol{Q}'\right)\boldsymbol{D}(\xi_n)\boldsymbol{Q}\left(\boldsymbol{Q}'\boldsymbol{D}(\xi_n)\boldsymbol{Q}\right)^{-1} \\ = \; & ch_{\max}\left(\boldsymbol{Q}'\boldsymbol{D}(\xi_n)\boldsymbol{Q}\right)^{-1}\boldsymbol{Q}'\boldsymbol{D}^2(\xi_n)\boldsymbol{D}(\xi_n)\boldsymbol{Q}\left(\boldsymbol{Q}'\boldsymbol{D}(\xi_n)\boldsymbol{Q}\right)^{-1} - 1. \end{aligned}$$

Here we use that the maximum eigenvalue of a matrix $\boldsymbol{P}\boldsymbol{P}'$ is that of $\boldsymbol{P}'\boldsymbol{P}$. Finally, and with

$$\boldsymbol{R}(\xi) = \boldsymbol{Q}'\boldsymbol{D}(\xi)\boldsymbol{Q},\ \boldsymbol{S}(\xi) = \boldsymbol{Q}'\boldsymbol{D}^2(\xi)\boldsymbol{Q},\ \boldsymbol{U}(\xi) = \boldsymbol{R}^{-1}(\xi)\boldsymbol{S}(\xi)\boldsymbol{R}^{-1}(\xi),$$

(11) becomes

$$I_\nu(\xi_n) = (1-\nu)\, tr\boldsymbol{R}^{-1}(\xi_n) + \nu ch_{\max}\boldsymbol{U}(\xi_n). \tag{12}$$

This is precisely the value minimized, in Wiens (2018), to obtain minimax design for LS estimates, thus justifying our statement in §1 that such designs remain minimax optimal, asymptotically, for M-estimation.

The minimization of $I_\nu(\xi_n)$ is carried out sequentially, as described in Theorem 5 of Wiens (2018). Briefly, given a current k-point design ξ_k, the loss resulting from the addition of a design point at $\boldsymbol{x}_i$ is expanded as

$$I_\nu\left(\xi_{k+1}^{(i)}\right) = I_\nu(\xi_k) - t_{k,i}/k + O\left(k^{-2}\right), \tag{13}$$

and then $\boldsymbol{x}_{(i)}$, with $(i) = \arg\max_i t_{k,i}$, is added to the design. This is carried out to convergence, yielding a design, with weights $\{\xi_{1,n}, \cdots, \xi_{N,n}\}$, on χ. Typically most $\xi_{i,N}$ are zero, but the remaining allocations $n\xi_{i,N}$ are not integers. To obtain implementable designs we first round up the n_i to $\lceil n\xi_{i,N}\rceil$, whose sum then exceeds n. The excess is decreased stepwise, by removing points whose value of $t_{n,i}$ in (13) is a minimum. This method typically results in only a very small increase in the minimized value of $I_\nu(\xi_n)$.

3. Examples and discussion

To obtain minimizing designs only ν, in (12), need be specified by the user. The simplest and most natural way to do this is to view ν as expressing the relative emphasis on the reduction of losses due to bias, versus those due to variation. Its choice is then up to the user; quite typically $\nu = .5$ is chosen. In this method there is NO difference in the minimax designs for different M-estimates.

Another method is suggested by the definition of ν at (9). Although the parameters involved in this definition would not be known to the designer, it is of interest to see how their values could affect the resulting designs.

We begin by seeing how much ν can change from its value under LS. The relationship between ν using Least Squares and ν using the M-estimate is, with $\gamma \overset{def}{=} \sigma/\kappa$, that

$$\nu_{\text{LS}} = \left(\gamma^2 + 1\right)^{-1} \text{ and } \nu_{\text{M}} = \left(\frac{\gamma^2\sigma_M^2}{\sigma^2} + 1\right)^{-1}.$$

We evaluate assuming that $\varepsilon/\sigma \sim N(0,1)$ and that $\psi(x) = xI(|x| \le c) + cI(|x| > c)$ (Huber (1964)).

Lemma 1. *With notation as above,*

$$0 \le \nu_{\text{LS}} - \nu_{\text{M}} \le \frac{\sqrt{\pi/2} - 1}{\sqrt{\pi/2} + 1} \simeq .1124.$$

The lower bound is attained only when the M-estimate is the LSE, *and the upper bound is attained when the M-estimate is the L_1 estimate and $\gamma^2 = 1/\sqrt{\pi/2} \simeq .7979$. At the maximum*

$$\nu_{\text{LS}} = \frac{\sqrt{\pi/2}}{\sqrt{\pi/2} + 1} \simeq .5562, \quad \nu_{\text{M}} = \frac{1}{\sqrt{\pi/2} + 1} = 1 - \nu_{\text{LS}} \simeq .4438. \tag{14}$$

See Figure 1.

We have constructed designs, using the 'worst case' values ν_{LS} and ν_{M} of Lemma 1 and also their midpoint $\nu = .5$. See Figure 2 for linear regression and Figure 3 for cubic regression, each on design spaces consisting of 20 equally spaced points spanning $[-1, 1]$. Given that ν_{LS}, ν_{M} and the M-estimate were chosen so as to make the designs as different as possible, these designs are remarkably similar – in some cases identical. The implementations all can be described as taking the replicates that would otherwise be assigned by the classically ($\nu = 0$) I-optimal designs and spreading them out into clusters at nearby design points. The I-optimal design for linear regression places mass of .5 at each of ± 1. That for cubic regression was derived by Studden (1977) and places masses of .1545 and .3455 at ± 1 and $\pm .4472$.

We conclude that an experimenter should feel quite safe in using the same design for an experiment regardless of which M-estimate is to be employed, and in choosing ν to represent his desired emphasis on bias reduction, as posited at the beginning of this section.

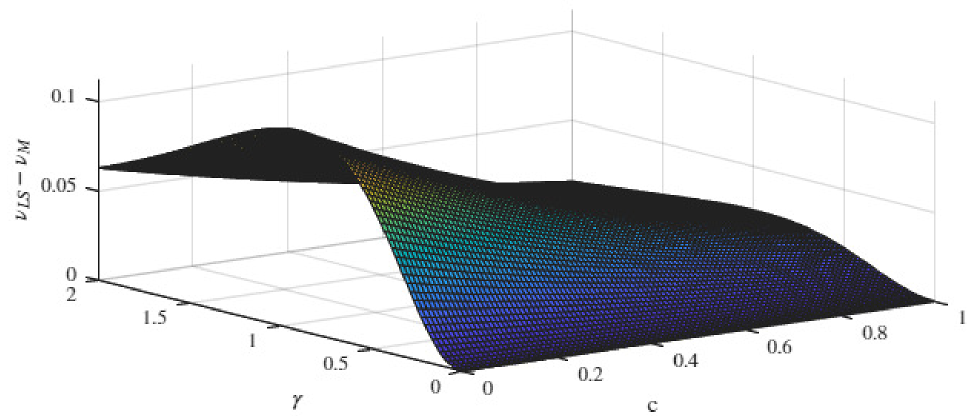

Figure 1: Differences $\nu_{LS} - \nu_M$ in terms of γ and c.

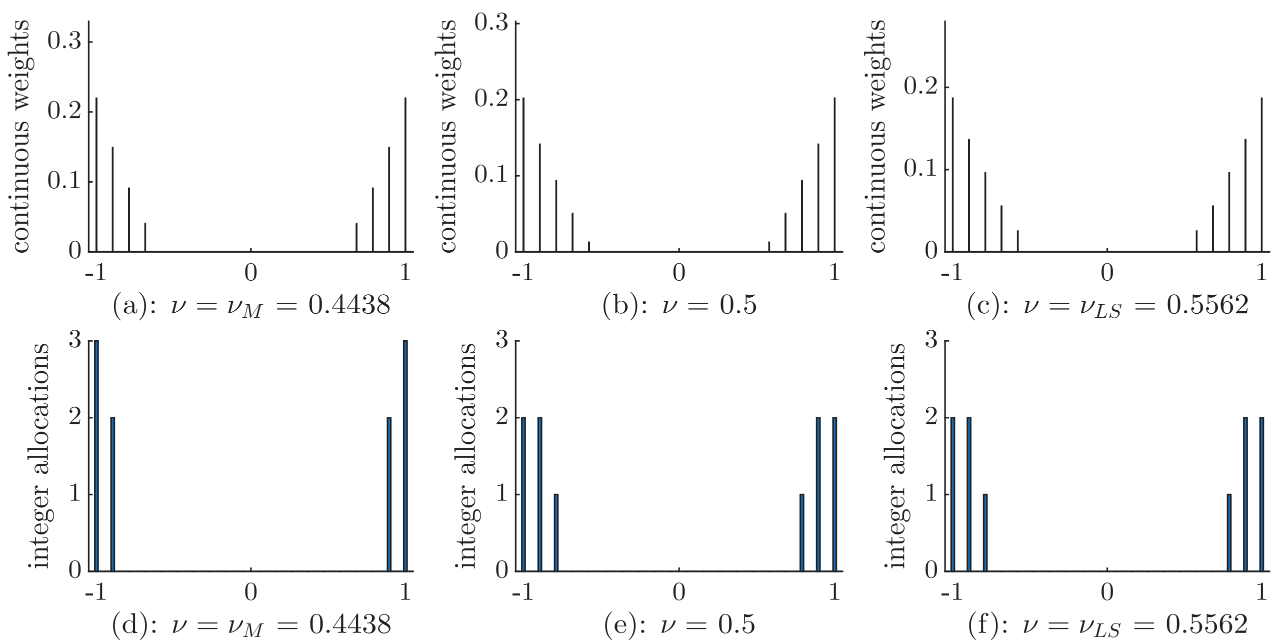

Figure 2: Designs for linear regression; n=10, N=20. (a)-(c): Continuous weights; (d)-(f): Integer allocations.

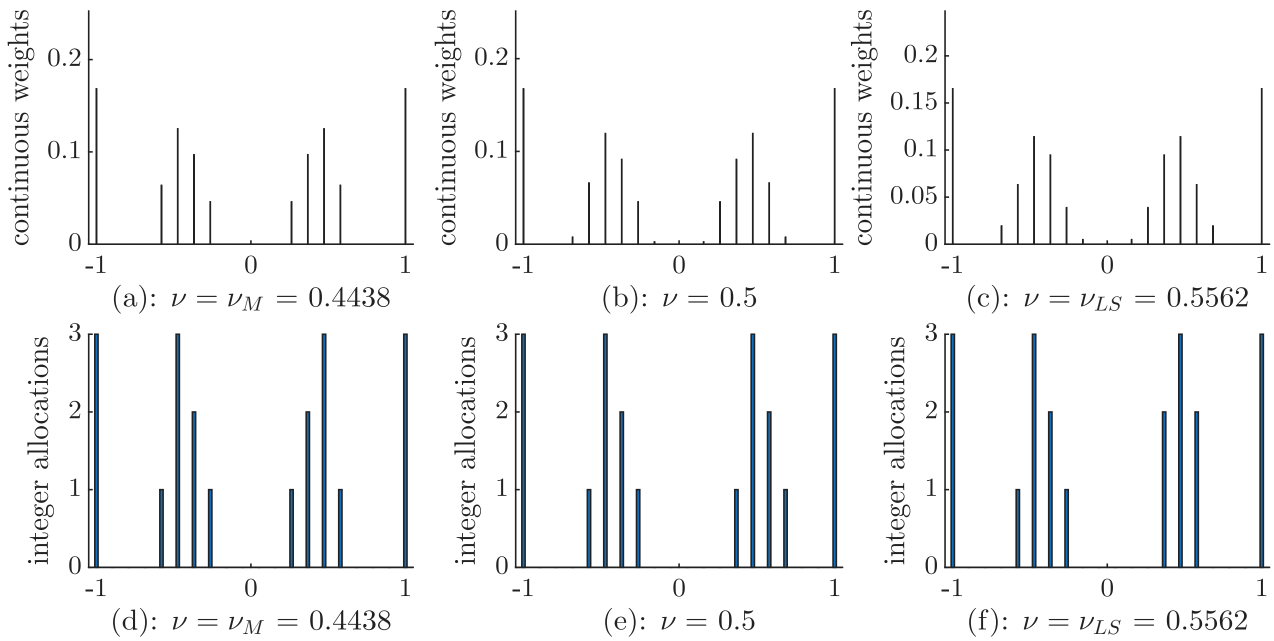

Figure 3: Designs for cubic regression; n=20, N=20. (a)-(c): Continuous weights; (d)-(f): Integer allocations.

Appendix: Proofs

A.1. Proof of Theorem 1

Still to come

A.2. Proof of Lemma 1

We calculate that

$$\frac{\sigma_M^2}{\sigma^2} = \frac{1 - 2c\phi(c) + 2\left(c^2 - 1\right)\Phi(-c)}{(1 - 2\Phi(-c))^2} \stackrel{def}{=} G(c),$$

in terms of which $\nu_{\text{M}} = \left(\gamma^2 G(c) + 1\right)^{-1}$. The function G is the asymptotic variance of an M-estimate of location under normality. It is decreasing in c and must exceed the asymptotic variance of the sample mean: $G(c) > G(\infty) = 1$. Two applications of l'Hopital's rule give $G(0) = \pi/2$, the asymptotic variance of the median of a normal sample, with corresponding regression estimate given by $\rho(x) = |x|$. Thus

$$0 < \nu_{\text{LS}} - \nu_{\text{M}} = \frac{\gamma^2}{\gamma^2 + 1}\left\{\frac{G(c) - 1}{\gamma^2 G(c) + 1}\right\} \stackrel{def}{=} H\left(c, \gamma^2\right).$$

For fixed γ, $H\left(c, \gamma^2\right)$ is increasing in $G(c)$, so is maximized at $c = 0$. Then $H\left(0, \gamma^2\right)$ vanishes at $\gamma^2 = 0, \infty$ and is maximized at $\gamma^2 = [G(0)]^{-1/2}$, with value $\max_{c,\gamma}(\nu_{\text{LS}} - \nu_{\text{M}}) = \left(\sqrt{G(0)} - 1\right) \Big/ \left(\sqrt{G(0)} + 1\right)$. At the maximum, $\nu_{\text{LS}} = 1/\left(\gamma^2 + 1\right)$ and $\nu_{\text{M}} = 1/\left(\gamma^2 G(0) + 1\right)$, giving (14) . □

Acknowledgements

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

References

Huber, P. J. (1964), "Robust Estimation of a Location Parameter," *The Annals of Mathematical Statistics*, 35, 73–101.

Rocke, D. and Shannon, D. (1986), "The Scale Problem in Robust Regression M-estimates," *Journal of Statistical Computation and Simulation*, 24, 47–69.

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

Wiens, D. P. (1994), "Robust designs for Approximately Linear Regression: M-Estimated Parameters," *Journal of Statistical Planning and Inference*, 40, 135–160.

— (1996), "Asymptotics of Generalized M-Estimation of Regression and Scale With Fixed Carriers, in an Approximately Linear Model," *Statistics and Probability Letters*, 30, 271–285.

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

Wiens, D. P. and Wu, E. K. H. (2010), "A Comparative Study of Robust Designs for M-Estimated Regression Models," *Computational Statistics and Data Analysis*, 54, 1683–1695.