A comparative study of robust designs for M-estimated regression models

Douglas P. Wiens\textsuperscript{a,}\textsuperscript{*}, Eden K.H. Wu\textsuperscript{b}

\textsuperscript{a} Department of Mathematical and Statistical Sciences, University of Alberta, Edmonton, Alberta, Canada T6G 2G1
\textsuperscript{b} Department of Statistics, Chinese University of Hong Kong, Hong Kong

\begin{abstract}
We obtain designs which are optimally robust against possibly misspecified regression models, assuming that the parameters are to be estimated by one of several types of M-estimation. Such designs minimize the maximum mean squared error of the predicted values, with the maximum taken over a class of departures from the fitted response function. One purpose of the study is to determine if, and how, the designs change in response to the robust methods of estimation as compared to classical least squares estimation. To this end, numerous examples are presented and discussed.
\end{abstract}

\section{Introduction}

The study of design strategies for experiments whose data are to be analyzed by least squares based regression methods, and which are to be robust against various model misspecifications, is well advanced (Box and Draper, 1959; Huber, 1975; Li and Notz, 1982; Wiens, 1992; Wiens and Zhou, 1997). Less is known about the situation in which M-estimation is to be the preferred regression approach. In this article we attempt to answer the following questions:

1. Is a design, optimally robust in anticipation of a least squares regression, still (approximately) so for M-estimation?
2. Common performance measures for a design depend only on the first two moments of the estimates, in the least squares case. For M-estimates, there are as well other quantities which depend on the error distribution. How do the designs change in response to changing error distributions, to changing types of score functions, or their tuning parameters?
3. Generalized ("Bounded Influence") M-estimation is a method which weights the covariates, as well as downweighting outliers in the response space. Given that aberrant leverage points are not anticipated to pose a problem in designed experiments, is there a role to be played by such weights at the design stage?

Brief answers to these questions, to be expanded upon in the sequel, appear to be: 1. Yes. 2. Little dependence on the error distributions; somewhat more on the types of score functions used (Huber, Mallows, Schweppes, redescenders, etc.). 3. The effect of the weights on the designs seems to be determined by a combination of the type of score function used and
the relative emphasis on variance versus bias. Regardless of this relative emphasis, the value of the mse is quite dependent on the scores and weights.

Practitioners wondering if they should use a robust method of estimation, rather than least squares, are sometimes told to compute both, and to use significant differences between the two as a possible warning. Based on the results reported here, we can offer analogous advice. For a preliminary study, a design tailored for least squares (but robust against model misspecifications of the form (2.2)) will generally be quite adequate, if sub-optimal. Once the experimenter has acquired a feeling for the anticipated nuances of the data, he is advised to compute a range of possible designs, from which to make a final choice in order to minimize the resulting mse.

2. Optimally robust regression designs

The first step in our study is to find designs for regression models, robust against model misspecification, which are tuned for use with M-estimated parameters. The framework is that we take \( n \) independent observations from an approximately linear regression model with independent, additive, homoscedastic errors:

\[
Y(x) = E[Y|x] + e.
\]

Here \( x \) is a \( d \)-dimensional independent variable ranging over a finite design space \( \mathcal{S} = \{x_1, \ldots, x_N\} \), and

\[
E[Y|x] \approx z^T(x) \beta,
\]

for a \( p \)-dimensional regressor \( z \). We define the “best” linear approximation via

\[
\tilde{\beta} = \arg\min_{\beta} \sum_{x_i} \{E[Y|x_i] - z^T(x_i) \beta\}^2,
\]

and set

\[
f_n(x) = \sqrt{n} \left\{ E[Y|x] - z^T(x) \tilde{\beta} \right\}.
\]

Then

\[
\sum_{x_i} z(x_i) f_n(x_i) = 0,
\]

and the observations are

\[
Y_i = z^T(x_i) \tilde{\beta} + n^{-1/2} f_n(x_i) + e_i.
\]

We require as well

\[
N^{-1} \sum_{x_i} f_n^2(x_i) \leq \eta^2,
\]

for a fixed constant \( \eta^2 \). A bound is necessary here in order that errors due to bias not swamp those due to variation. The inclusion of the \( \sqrt{n} \) in the definition of \( f_n \) ensures, via (2.3), that these types of error are of the same order asymptotically; it is analogous to the notion of contiguity in the asymptotic theory of hypothesis testing.

By a design we mean any probability distribution \( \{p_i\}_{i=1}^N \) on \( \mathcal{S} \); if \( p_i = n_i/n \) for integers \( n_i \geq 0 \) we say that the design is exact. An exact design is implemented by allocating \( n_i \) observations to \( x_i \), resulting in observations \( \{y_i\}_{i=1}^n \). Only exact designs are considered here.

Suppose that \( \tilde{\theta}_n = t(\tilde{\beta}_n, \tilde{\sigma}_n) \) is an M-estimate of regression/scale, satisfying

\[
\frac{1}{n} \sum_{i=1}^n \Omega_i(\tilde{\theta}_n) = o_p(n^{-1/2}),
\]

where

\[
\Omega_i(\theta) = \begin{pmatrix} \psi_i \left( \frac{Y_i - z^T(x_i) \beta}{\sigma} \right) z(x_i) \\ \chi_i \left( \frac{Y_i - z^T(x_i) \beta}{\sigma} \right) - A_n \end{pmatrix}
\]

for a weakly increasing, absolutely continuous, bounded function \( \psi_i \),

\[
\chi_i(r) = \int_0^r t \psi'_i(t) \, dt,
\]
and a bounded sequence of positive constants \( \{A_n\} \). Examples are furnished by the Generalized M-estimators, with

\[
\psi_i(r) = w(\mathbf{x}_i) \psi \left( \frac{r}{v(\mathbf{x}_i)} \right).
\]

We shall employ either \( v(\mathbf{x}_i) \equiv 1 \), giving a Mallows GM-estimate (Hill, 1977) or \( v(\mathbf{x}_i) = w(\mathbf{x}_i) \), giving a Schweppe estimate (Merrill and Schweppe, 1971). Huber (“Ordinary”) M-estimates are Mallows GM-estimates with constant weights \( w(\mathbf{x}_i) \equiv 1 \). In our examples we use weights \( w_k(\mathbf{x}_i) = (1 + k\|\mathbf{x}_i\|^2)^{-\frac{1}{2}} \), so that Ordinary M-estimates are Mallows estimates with \( k = 0 \). From (2.5) the score function for the scale estimate is, with \( w_i \equiv w(\mathbf{x}_i) \) and \( \chi(r) \equiv \int_0^r t \psi'(t) \, dt \), given by

\[
\chi_i(r) = \begin{cases} 
   w_i \chi(r), & \text{for Mallows estimates,} \\
   w_i^2 \chi \left( \frac{r}{w_i} \right), & \text{for Schweppe estimates.}
\end{cases}
\]

In Wiens (1996), conditions are given under which \( \hat{\theta}_n \) is consistent for \( \theta = (\hat{\beta}, \sigma_n) \), where \( \sigma_n \) is a sequence of positive constants satisfying

\[
\sum_{i=1}^N p_i E \left[ \psi_i \left( \frac{e}{\sigma_n} \right) \right] z(\mathbf{x}_i) = o \left( n^{-1/2} \right),
\]

\[
\sum_{i=1}^N p_i E \left[ \chi_i \left( \frac{e}{\sigma_n} \right) - A_n \right] = o \left( n^{-1/2} \right),
\]

(2.6)

and is asymptotically normal:

\[
\sqrt{n} \left( \hat{\theta}_n - \theta \right) \sim AN \left( M_n^{-1} b_n, M_n^{-1} Q_n M_n^{-1} \right),
\]

for

\[
b_n = \frac{1}{\sigma_n} \sum_{i=1}^N p_i \left( E \left[ \psi_i \left( \frac{e}{\sigma_n} \right) \right] z(\mathbf{x}_i) f_n(\mathbf{x}_i) \right),
\]

\[
M_n = \frac{1}{\sigma_n^2} \sum_{i=1}^N p_i \left( E \left[ \psi_i^2 \left( \frac{e}{\sigma_n} \right) \right] z(\mathbf{x}_i) z^T(\mathbf{x}_i) \right),
\]

\[
Q_n = \sum_{i=1}^N p_i \left( E \left[ \psi_i^2 \left( \frac{e}{\sigma_n} \right) \right] z(\mathbf{x}_i) z^T(\mathbf{x}_i) \right).
\]
Fig. 2. Minimax designs for straight line regression using GM-estimation and varying scores $\psi_i$. All cases use weights $w_{10}(x), \varepsilon = 0.25$ and $\pi = 0.5$.

Upper panel: Mallows weighting. (a) $c = 0.5$ ($\xi = 20.31, \mu_2 = 0.87, \sigma_n = 1.06$), (b) $c = 1.5$ ($\xi = 17.62, \mu_2 = 0.82, \sigma_n = 1.19$), (c) $c = 5$ ($\xi = 16.78, \mu_2 = 0.82, \sigma_n = 1.87$).

Lower panel: Schweppe weighting. (a) $c = 0.5$ ($\xi = 22.84, \mu_2 = 0.87, \sigma_n = 1.05$), (b) $c = 1.5$ ($\xi = 20.26, \mu_2 = 0.82, \sigma_n = 1.06$), (c) $c = 5$ ($\xi = 16.27, \mu_2 = 0.63, \sigma_n = 1.35$).

Let $Z$ be the $N \times p$ matrix with rows $\{z^i(x_i)\}_{i=1}^N$, $D_1$ and $D_2$ the $N \times N$ diagonal matrices with diagonal elements

$$d_{1,ii} = E \left[ \psi_i \left( \frac{e}{\sigma_n} \right) \right] = \begin{cases} \frac{w_i E \left[ \psi_i \left( \frac{e}{\sigma_n} \right) \right]}{E \left[ \psi_i \left( \frac{e}{\sigma_n} \right) \right]}, & \text{for Mallows estimates,} \\ \frac{w_i E \left[ \psi_i \left( \frac{e}{\sigma_n} \right) \right]}{E \left[ \psi_i \left( \frac{e}{\sigma_n} \right) \right]}, & \text{for Schweppe estimates;} \end{cases}$$

(2.7a)

$$d_{2,ii} = E \left[ \psi_i^2 \left( \frac{e}{\sigma_n} \right) \right] = \begin{cases} \frac{w_i^2 E \left[ \psi_i \left( \frac{e}{\sigma_n} \right) \right]}{E \left[ \psi_i \left( \frac{e}{\sigma_n} \right) \right]}, & \text{for Mallows estimates,} \\ \frac{w_i^2 E \left[ \psi_i \left( \frac{e}{\sigma_n} \right) \right]}{E \left[ \psi_i \left( \frac{e}{\sigma_n} \right) \right]}, & \text{for Schweppe estimates;} \end{cases}$$

(2.7b)

Let $f$ be the $N \times 1$ vector with elements $\{f_i(x_i)\}_{i=1}^N$ and $P$ the diagonal matrix with diagonal elements $\{p_i = n_i/n\}_{i=1}^N$ (note that some of the $n_i$ will typically be zero). Assume that $Z$ has full column rank, that $\psi$ is an odd function of its argument, and that the errors are symmetrically distributed. Then the regression and scale components of $\hat{\beta}_n$ are asymptotically independent, and

$$\sqrt{n} \left( \hat{\beta}_n - \bar{\beta} \right) \sim AN \left( \left( Z^T D_1 P Z \right)^{-1} Z^T D_1 P f, \sigma_n^2 \left( Z^T D_1 P Z \right)^{-1} \left( Z^T D_2 P Z \right) \left( Z^T D_1 P Z \right)^{-1} \right).$$

The asymptotic Average Mean Squared Error (AMSE) of $\hat{Y}_n(x) = z^i(x)\hat{\beta}_n$ as an estimate of $E[Y|x] = z^i(x)\bar{\beta} + n^{-1/2} f_n(x_i)$ is, apart from terms which are $o(1)$,

$$\text{AMSE} = \frac{1}{N} \sum_{i=1}^N E \left[ \left( \sqrt{n} \left( \hat{Y}_n(x_i) - E[Y|x_i] \right) \right)^2 \right]$$

$$= \frac{1}{N} \text{tr} \left( \text{cov} \left[ \sqrt{n} \left( \hat{\beta}_n - \bar{\beta} \right) \right] Z^T Z \right) + \frac{1}{N} \left\| Z E \left[ \sqrt{n} \left( \hat{\beta}_n - \bar{\beta} \right) \right] \right\|^2 + \frac{1}{N} \sum_{i=1}^N f_i^2(x_i)$$

$$= \frac{\sigma_n^2}{N} \text{tr} \left( \left( Z^T D_1 P Z \right)^{-1} \left( Z^T D_2 P Z \right) \left( Z^T D_1 P Z \right)^{-1} Z^T Z \right) + \frac{1}{N} \left\| Z \left( Z^T D_1 P Z \right)^{-1} Z^T D_1 P f \right\|^2 + \frac{1}{N} \| f \|^2.$$
We maximize AMSE over \( f \), subject to (2.1) and (2.3), and then minimize the result over probability distributions \( \{p_i = n_i/n\} \) on \( \delta \), thus obtaining a minimax exact design. This first step is carried out analytically, in a manner similar to that in Fang and Wiens (2000), and is briefly described as follows. Let \( Z = U_{n \times p} A_{p \times p} V_{p \times p}^T \) be the singular value decomposition, with \( U^T U = V^T V = I_p \) and \( A \) diagonal and invertible. Augment \( U \) by \( \tilde{U}_{N \times N-p} \) in such a way that \( U^T \tilde{U} \) is orthogonal. Then by (2.1) and (2.3) we have that there is an \( N - p \times 1 \) vector \( c \), with \( \|c\| \leq 1 \), satisfying

\[
 f = \eta / \sqrt{N} \tilde{U} c.
\]

With \( c_{\max} \) denoting a maximum eigenvalue, the maximum AMSE becomes

\[
 \max \text{AMSE} = \frac{\sigma_n^2}{N} \text{tr} \left\{ (U^T D_1 P U)^{-2} (U^T D_2 P U) \right\} + \eta^2 \left\{ 1 + c_{\max} U^T P D_1 U (U^T D_1 P U)^{-2} U^T D_1 P U \right\}. \tag{2.8}
\]

Since \( \tilde{U} \tilde{U}^T = I_N - U U^T \) we obtain

\[
 c_{\max} U^T P D_1 U (U^T D_1 P U)^{-2} U^T D_1 P U = c_{\max} \left\{ (U^T D_1 P U)^{-2} (U^T D_1 P^2 D_1 U) \right\} - 1; \tag{2.9}
\]

this requires only the eigenvalues of a \( p \times p \) matrix. Upon substituting (2.9) into (2.8), and defining

\[
 \pi = \frac{\sigma_n^2}{N} + \eta^2 \in [0, 1],
\]

we have that \( \max \pi, \text{AMSE} \) is given by \( \sigma_n^2 / N + \eta^2 \) times

\[
 \mathcal{L}_\pi = \pi \cdot \text{tr} \left\{ (U^T D_1 P U)^{-2} (U^T D_1 P U) \right\} + (1 - \pi) c_{\max} \left\{ (U^T D_1 P U)^{-2} (U^T D_1 P^2 D_1 U) \right\}. \tag{2.10}
\]

The problem is now to minimize \( \mathcal{L}_\pi \) for fixed \( \pi \). Note that the experimenter is free to choose \( \pi \) according to the relative importance, assigned by him, to bias versus variance—\( \pi = 0 \) corresponds to a “pure bias” problem, \( \pi = 1 \) to “pure variance”. 

---

*Fig. 3.* Minimax designs for straight line regression using GM-estimation and varying weights \( w_k(x) \) when bias is dominant: \( \pi = 0.05 \). All cases use scores \( \psi_{1.5}(r) \) and \( \delta = 0.25 \). Upper panel: Mallows weighting. (a) \( k = 0 \) (\( \lambda = 3.10, \mu_2 = 0.48 \)), (b) \( k = 10 \) (\( \lambda = 3.23, \mu_2 = 0.56 \)), (c) \( k = 20 \) (\( \lambda = 3.30, \mu_2 = 0.56 \)). All have \( \sigma_n = 1.19 \). Lower panel: Scheppe weighting. (a) \( k = 0 \) (\( \lambda = 3.09, \mu_2 = 0.48, \sigma_n = 1.17 \)), (b) \( k = 10 \) (\( \lambda = 3.18, \mu_2 = 0.48, \sigma_n = 1.12 \)), (c) \( k = 20 \) (\( \lambda = 3.33, \mu_2 = 0.55, \sigma_n = 1.11 \)).
Fig. 4. Minimax designs for straight line regression using GM-estimation and varying weights \( w_k(x) \) when variance is dominant: \( \pi = 0.85 \). All cases use scores \( \psi_{1,5}(r) \) and \( \varepsilon = 0.25 \). Upper panel: Mallows weighting. (a) \( k = 0 \) (\( L = 26.82, \mu_2 = 0.94 \)), (b) \( k = 10 \) (\( L = 26.84, \mu_2 = 0.96 \)), (c) \( k = 20 \) (\( L = 26.84, \mu_2 = 0.96 \)). All have \( \sigma_n = 1.17 \). Lower panel: Schweppeweighting. (a) \( k = 0 \) (\( L = 31.74, \mu_2 = 0.94, \sigma_n = 1.06 \)), (b) \( k = 10 \) (\( L = 33.16, \mu_2 = 0.96, \sigma_n = 1.06 \)).

We minimize (2.10) over exact designs through simulated annealing—see e.g. Fang and Wiens (2000). The method requires values of the elements of \( D_1 \) and \( D_2 \) at all \( x_i \). So we assume that the errors follow a particular distribution—Normal, or Normal with some contamination—and express \( \sigma_n \) as a functional of this error distribution. From this we calculate the terms in (2.7) and then (2.10).

For Fisher consistency at the Normal distribution, in (2.4) we employ constants \( A_n = \sum_{i=1}^{N} p_i E_\phi[\chi_i(e)] \). We take the error distribution \( F \) to be an \( \varepsilon \)-contaminated Normal distribution:

\[
F = (1-\varepsilon) \Phi + \varepsilon G,
\]

for a symmetric but otherwise arbitrary distribution function \( G \). Then (2.6) yields, up to terms which are \( o(n^{-1/2}) \),

\[
E = \frac{\sum_{i=1}^{N} p_i \left\{ E_\Phi[\chi_i(e)] - E_\phi[\chi_i\left(\frac{e}{\sigma_n}\right)] \right\}}{\sum_{i=1}^{N} p_i \left\{ E_G[\chi_i\left(\frac{e}{\sigma_n}\right)] - E_\phi[\chi_i\left(\frac{e}{\sigma_n}\right)] \right\}},
\]

\[
E = \begin{cases} 
E_\phi[\chi(e)] - E_\phi[\chi\left(\frac{e}{\sigma_n}\right)], & \text{for Mallows estimates,} \\
E_G[\chi\left(\frac{e}{\sigma_n}\right)] - E_\phi[\chi\left(\frac{e}{\sigma_n}\right)], & \text{for Schweppew estimates.}
\end{cases}
\]

(2.11)

The choices of \( G \) investigated are (i) \( t_1 \), the Student’s \( t \) on 1 degree of freedom (i.e. the Cauchy), (ii) \( t_3 \), a popular choice in robustness studies since it has rather long tails but is not as extreme as the Cauchy, and (iii) \( N_9 \), Normal with a variance of 9. For these choices the expectations in (2.7) and (2.11) can be evaluated explicitly. In most cases we display the results only for the Cauchy, the other cases being very similar (see Figs. 7 and 8 for illustrations of this).
Fig. 5. Optimal designs for straight line regression using a GM-estimator and varying emphasis $\pi$ on variance versus bias. All cases use scores $\psi_{1.5}(r)$, $\varepsilon = 0.25$ and weights $w_{10}(x)$. Upper panel: Mallows weights. Lower panel: Schweppes weighting. (a) $\pi = 0$, (b) $\pi = 0.25$, (c) $\pi = 0.5$, (d) $\pi = 0.75$, (e) $\pi = 1.0$.

The annealing step is as follows. Given a current design on $n$ points, we randomly choose one of the sampled locations and remove a point from it. We then randomly choose one of the $N$ locations in $S$ and have that observation assigned to it. The loss is then recalculated. Let $\nabla L$ denote the increase in the loss resulting from this change in design. Then the new design is accepted if

$$e^{-\nabla L/T} > U,$$

where $U$ is a simulated Unif(0, 1) r.v. Thus a superior design ($\nabla L < 0$) is accepted with probability 1, an inferior one with probability $e^{-\nabla L/T}$. The “temperature” parameter $T$ is chosen in such a way that, initially, about 50% of all states are accepted; it is then decreased by a factor of 0.9 after each 100 trials. See Bohachevsky et al. (1986) and Haines (1987) for background material.

All computations were carried out in MATLAB, and the code is available from the authors.

3. One-dimensional Mallows GM-estimation

As a first step we implement the development of Section 2 for straight line regression on $[-1, 1]$, using Huber (1964) scores $(\psi_c(r) = \psi(r) = \max(-c, \min(r, c)))$, for which $\chi(r) = \psi^2(r)/2$, and Mallows weights $w_k(x)$, obtaining $\psi_i(r) = w_k(x_i)\psi(r)$. To evaluate (2.7) and (2.11) we first define

$$I_F(a, b) = \int_0^a z^b dF(z) \quad \text{and} \quad H(c; F) = \frac{I_F(c, 2)}{c^2} + F(-c),$$

then with $c_n \equiv c\sigma_n$ we calculate that

$$E_F[\chi(e)] = c^2 H(c; F), \quad E_F\left[\chi\left(\frac{e}{\sigma_n}\right)\right] = c^2 H(c_n; F),$$

$$E_F\left[\psi'(\frac{e}{\sigma_n})\right] = 1 - 2F(-c_n), \quad E_F\left[\psi^2\left(\frac{e}{\sigma_n}\right)\right] = 2c^2 H(c_n; F).$$

In preparing the examples only symmetric designs were admitted for consideration in the annealing process, and 4000 trials were run. Repeated runs resulted in the same designs, suggesting that a global optimum was found. The examples use
Fig. 1.

Fig. 2.

Fig. 3.

Fig. 4.

Fig. 5.

Fig. 6. Minimax designs for cubic regression using GM-estimation, scores $\psi_{1,5}(r)$, weights $w_{i0}(x)$, $\epsilon = 0.15$ and varying $\pi$. Upper panel: Mallows weighting. (a) $\pi = 0.25 (\mathcal{L} = 19.39, \mu_2 = 0.46, \sigma_1 = 1.06)$, (b) $\pi = 0.5 (\mathcal{L} = 36.46, \mu_2 = 0.44, \sigma_1 = 1.07)$. Lower panel: Schweppe weighting. (a) $\pi = 0.25 (\mathcal{L} = 21.47, \mu_2 = 0.53, \sigma_1 = 1.04)$, (b) $\pi = 0.5 (\mathcal{L} = 40.19, \mu_2 = 0.51, \sigma_1 = 1.05)$.

$n = N = 20, S = \{\pm 0.1, \pm 0.2, \ldots, \pm 1.0\}$ and $G =$ Cauchy, but other choices give the same message. In the captions of the figures the second moments $\sum_{i=1}^{N} p_i x_i^2$ of the designs are denoted by $\mu_2$.

As a benchmark we exhibit, in Fig. 1, the designs which are minimax for least squares estimation, i.e. $\epsilon = k = 0, c = \infty$. See (the top panel of) Fig. 2 for plots of the designs for Mallows GM-estimation using $c = 0.5, 1.5, 5$ with $\epsilon = 0.25, k = 10, \pi = 0.5$. The designs turn out to be identical for the two larger values of $c$. Figs. 3 and 4 illustrate examples of designs as the weights vary, in each case using $c = 1.5, \epsilon = 0.25$. In Fig. 3 we took $\pi = 0.05$, while in Fig. 4, $\pi = 0.85$. In both cases there is a tendency for the mass to migrate to the extremes of the design space as $k$ increases. In both cases the mse $\mathcal{L}_n$ increases, albeit only slightly, as $k$ increases. There is of course a dependence on $\pi$—see Fig. 5, in which $\pi = 0(0.25)1$ with $c = 1.5, \epsilon = 0.25, k = 10$. Fig. 6 gives comparative designs for cubic regression.

4. One-dimensional Schweppe GM-estimation

For a Schweppe estimate based on Huber’s $\psi_c$, the evaluations of (2.7) and (2.11) utilize

$$E_F \left[ \chi \left( \frac{e}{w_i} \right) \right] = c^2 H(c w_i; F), \quad E_F \left[ \chi \left( \frac{e}{w_i \sigma_n} \right) \right] = c^2 H(c_n w_i; F),$$

$$E_F \left[ \psi' \left( \frac{e}{w_i \sigma_n} \right) \right] = 1 - 2F(-c_n w_i), \quad E_F \left[ \psi^2 \left( \frac{e}{w_i \sigma_n} \right) \right] = 2c^2 H(c_n w_i; F).$$

Corresponding to the situations pictured in the top panels of Figs. 2–5, the designs pictured in the bottom panels are close – sometimes identical – to those using Mallows estimation. There is a tendency for the designs based on Schweppe estimation to move mass into the centre of the design space as $c$ or $k$ increases. This is a tendency which in general imparts robustness against nonlinearity (in the response function) to the designs.

We note that the D-optimal designs, for least squares estimation of a cubic response when $\epsilon = 0, \pi = 1$, place one quarter of the design points at each of $\pm 0.447$ and $\pm 1$—see Pukelsheim (1993, p. 218). The robust designs for M-estimation can be roughly described as replacing these atoms with clusters of points at nearby locations, with the clusters becoming more varied as $\pi$ decreases, and in a manner somewhat dependent on the type of M-estimate being employed—see Fig. 6 for an illustration of this.
5. Redescending M-estimates

Note that Huber’s $\psi$-function gives constant but non-zero weights to extreme values. Practitioners often opt instead for scores which cut the influence of such points to zero, by using a “redescending” $\psi$-function such as Tukey’s biweight:

$$\psi_{biw}(r) = r \left[1 - (r/c)^2\right] I_{[-c,c]}(r).$$

In this section we repeat the analyses of Sections 3 and 4 using (5.1). In terms of

$$B_1(c; F) = \frac{I_{r}(c, 2)}{c^2} - \frac{3I_{r}(c, 4)}{c^4} + \frac{5I_{r}(c, 6)}{3c^6},$$

$$B_2(c; F) = \frac{6I_{r}(c, 2)}{c^2} - \frac{5I_{r}(c, 4)}{c^4},$$

$$B_3(c; F) = \frac{I_{r}(c, 2)}{c^2} - \frac{4I_{r}(c, 4)}{c^4} + \frac{6I_{r}(c, 6)}{c^6} - \frac{4I_{r}(c, 8)}{c^8} + \frac{I_{r}(c, 10)}{c^{10}},$$

we find that the relevant expectations for Mallows estimation are

$$E_F \left[ \chi \left( \frac{e}{\sigma_n} \right) \right] = c^2 B_1(c; F), \quad E_F \left[ \chi' \left( \frac{e}{\sigma_n} \right) \right] = c^2 B_1(c; F),$$

$$E_F \left[ \psi' \left( \frac{e}{\sigma_n} \right) \right] = 1 - 2F(-c_n) - 2B_2(c_n; F), \quad E_F \left[ \psi^2 \left( \frac{e}{\sigma_n} \right) \right] = 2c^2 B_3(c; F).$$

For Schweppe estimation these are

$$E_F \left[ \chi \left( \frac{e}{w_i} \right) \right] = c^2 B_1(c w_i; F), \quad E_F \left[ \chi \left( \frac{e}{w_i\sigma_n} \right) \right] = c^2 B_1(c w_i; F),$$

$$E_F \left[ \psi' \left( \frac{e}{w_i\sigma_n} \right) \right] = 1 - 2F(-c_n w_i) - 2B_2(c_n w_i; F), \quad E_F \left[ \psi^2 \left( \frac{e}{w_i\sigma_n} \right) \right] = 2c^2 B_3(c_n w_i; F).$$
Fig. 8. Minimax designs for straight line regression using Mallows GM-estimation with Tukey’s biweight $\psi_{би}(\tau), \epsilon = 0.25$ and varying error distributions. Upper panel: $\pi = 0.05$ (bias dominant). (a) t1 ($\mathcal{L} = 3.93, \mu_2 = 0.62, \sigma_0 = 0.88$), (b) t3 ($\mathcal{L} = 3.30, \mu_2 = 0.56, \sigma_0 = 1.23$), (c) N9 ($\mathcal{L} = 2.76, \mu_2 = 0.56, \sigma_0 = 1.26$). Lower panel: $\pi = 0.85$ (variance dominant). (a) t1 ($\mathcal{L} = 36.16, \mu_2 = 0.98, \sigma_0 = 0.88$), (b) t3 ($\mathcal{L} = 27.90, \mu_2 = 0.96, \sigma_0 = 0.98$), (c) N9 ($\mathcal{L} = 20.77, \mu_2 = 0.94, \sigma_0 = 1.26$).

It is commonly acknowledged that a redescender should not descend too quickly (due to the terms $d_l$, appearing in various denominators) and so we take $c = 5$ in our applications of the biweight. In Figs. 7 and 8 we vary the error distribution, using Mallows weighting and either $\epsilon \in \{0.05, 0.35\}$ (Fig. 7) or $\pi \in \{0.05, 0.85\}$ (Fig. 8). In each cases the effect of the error distribution is quite slight. See also Fig. 9, for cubic regression using biweights — compared to the designs using Mallows weighting, those using Schweppe weighting tend to place more mass towards $\pm 1$ and spread out the remaining mass more uniformly in regions near the D-optimal points $\pm 0.447$.

6. Two-dimensional optimal designs

In this section, we exhibit optimal designs for multiple linear regression on a grid $\mathbf{x} = (x_1, x_2)^T, x_1, x_2 \in (-1, 0.2)1$, with restrictions of symmetry and exchangeability imposed and with $\mathbf{x} = (0, 0)$ omitted from the design space; thus $N = 120$. We take $n = 80$. Figs. 10 and 12 provide the designs using Huber’s $\psi_1, 1.5(r)$ and various values of $k$, with Mallows and Schweppe estimates respectively. Increasing $k$ results in mass being moved towards the centre of the design space. When a redescender is used – Figs. 11 and 13 – the same conclusions are drawn. There is some dependence on the type of estimate used – Mallows or Schweppe – but (plots not shown) very little on the form of the contaminating distribution.

7. Summary and conclusions

We have constructed designs which are tuned for a combination of robustness and efficiency, when used in conjunction with M-estimation. The lack of explicit, exact expressions for the objects of the optimization methods preclude general conclusions valid in all cases. However, some general observations can be made. Some have been alluded to already; here we expand on and augment these conclusions.

- In the examples considered, the optimal designs depart at least slightly from those optimal for least squares estimation (LSE), in manners unique to the alternate method of estimation being employed.
- The tuning constant “$c$”, upon which interest generally focusses in discussions of the efficiency of the M-estimator, plays a role in parameterizing the designs when a Huber $\psi_c$ is used. The loss tends to decrease with increasing $c$ and decreasing
Fig. 9. Minimax designs for cubic regression using GM-estimation, scores $\psi_{bi}(r)$, weights $w_{10}(x)$, $\varepsilon = 0.25$ and varying $\pi$. Upper panel: Mallows weighting. (a) $\pi = 0.2$ ($L = 21.36, \mu_2 = 0.46, \sigma_n = 0.88$), (b) $\pi = 0.8$ ($L = 77.68, \mu_2 = 0.44, \sigma_n = 0.88$). Lower panel: Schweppe weighting. (a) $\pi = 0.2$ ($L = 16.59, \mu_2 = 0.77, \sigma_n = 1.49$), (b) $\pi = 0.8$ ($L = 58.35, \mu_2 = 0.55, \sigma_n = 1.36$).

Fig. 10. Minimax designs for Mallows GM-estimation and $\pi = 0.15$ with Huber $\psi_{1.5}$ and $\varepsilon = 0.15$. (a) $k = 0$ ($L = 52.03$), (b) $k = 10$ ($L = 74.71$), (c) $k = 20$ ($L = 84.44$). All have $\sigma_n = 1.07$.

Fig. 11. Minimax designs for Mallows GM-estimation and $\pi = 0.15$ with biweight $\psi_{bi}(r)$. (a) $k = 0$ ($L = 65.57$), (b) $k = 10$ ($L = 75.39$), (c) $k = 20$ ($L = 91.23$). All have $\sigma_n = 0.94$. 
Fig. 12. Minimax designs for Schweppe GM-estimation and $\pi = 0.15$ with Huber $\psi_{1.5}$. (a) $k = 0 (\mathcal{L} = 53.01, \sigma_n = 1.075)$, (b) $k = 10 (\mathcal{L} = 64.45, \sigma_n = 1.05)$, (c) $k = 20 (\mathcal{L} = 82.75, \sigma_n = 1.05)$.

Fig. 13. Minimax designs for Schweppe GM-estimation and $\pi = 0.15$ with biweight $\psi_{biw}$, (a) $k = 0 (\mathcal{L} = 62.26, \sigma_n = 0.94)$, (b) $k = 10 (\mathcal{L} = 84.64, \sigma_n = 1.11)$, (c) $k = 20 (\mathcal{L} = 86.25, \sigma_n = 1.21)$.

$k$, i.e. as the estimation method moves towards least squares. When the scores are redescending, the choice of $c$ can be a delicate matter which should probably still be considered independently of the design.

- As the relative importance $\pi$ of variance versus bias decreases, so that the demands for robustness against response misspecifications outweigh those for efficiency, the optimal designs place less mass at extreme (boundary) points and more in the central region of the design space. Indeed, as with LSE, this parameter $\pi$ seems to be the major factor regulating the various designs.

- For the univariate cases considered, when the weights $w_k(x)$, downweighting carriers near the boundaries of the design space, have an effect it tends to be that the optimal designs place more mass near these boundaries as the downweighting increases, when bias is dominant. When variance is dominant the situation is reversed. This is intuitively plausible — when the design is tailored for efficiency it may avoid placing covariates in those “extreme” locations where they will be downweighted. In the bivariate cases the situation is less clear.

In summary it appears that the need for, and contribution of, a design which is robust against departures from the assumed linear model, do not depend greatly upon the method of estimation employed. The optimally robust LSE-based designs can safely be used as starting designs, but for maximum efficiency and robustness should be refined after analyzing preliminary studies.

Acknowledgements

The research of Douglas Wiens is supported by the Natural Sciences and Engineering Research Council of Canada and that of Eden Wu is supported by a Hong Kong RGC Earmarked Research Grant. We thank the anonymous reviewers for their insightful comments leading to an improved article.

References


