Minimum Variance Designs
With Constrained Maximum Bias

Douglas P. Wiens[a,*]

[a]*Mathematical & Statistical Sciences, University of Alberta, Edmonton, Canada, T6G 2G1*

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. This mean squared error decomposes into a term arising solely from variation, and a bias term arising from the model errors. Here we consider the problem of designing so as to minimize the variance of the predictors, subject to a bound on the maximum (over model misspecifications) bias. We consider as well designing so as to minimize the maximum bias, subject to a bound on the variance. We show that solutions to both problems are given by the minimax designs, with appropriately chosen values of their tuning constants. Conversely, any minimax design solves each problem for an appropriate choice of the bound on the maximum bias or on the variance.

Keywords: coefficient of maximum bias, I-optimality, misspecified model, regression design, robust, sample size monotonicity.
2010 MSC: Primary 62F35, Secondary 62K05

1. Introduction and summary

An experimental design to fit a particular model is *robust* if its performance is stable under perturbations of the true model. The theory of robustness of design was largely initiated by Box and Draper (1959), who investigated the robustness of some classical experimental designs in the presence of certain model inadequacies, e.g. designs optimal for a low order polynomial response when the true response was a polynomial of higher order. Huber (1975) derived designs for straight line regression, robust in the presence of alternate response functions. Wiens (1990, 1992) extended these results to multiple regression responses and in a variety of other directions – see Wiens (2015) for a summary of these and other approaches to robustness of design.

Designs which are *minimax* in the presence of model misspecifications aim to minimize the maximum, over classes of alternate response models, of the integrated mean squared error (IMSE) of the predicted values. In this note we first briefly review the theory

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

of such designs, and discuss a decomposition of the IMSE into a convex combination of two terms – one arising solely from variation, and the other arising from the bias due to the model errors. It is often the case that designs which yield a small value of one of these terms do poorly with respect to the other. For instance the designs optimal with respect to the variance-based alphabetic optimality criteria concentrate their mass at a minimal number of points, and thus fare poorly, with large biases, in the presence of model misspecifications. On the other hand uniform-like designs reduce the bias while increasing the variance. Thus in §3 we propose two associated problems: (i) design so as to minimize the integrated variance of the predictors, subject to a bound on the maximum (over model misspecifications) bias, and (ii) design so as to minimize the maximum bias, subject to a bound on the variance. We show that solutions to both problems are given by the minimax designs, for appropriately chosen values of the mixing parameter. Conversely, any minimax design solves both problems for appropriate choices of the bounds on the maximum bias or variance.

Examples and methods of implementation are studied in §4. The MATLAB code used to prepare these examples is available on the author's personal website.

2. Minimax robustness of design

The general 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 $\boldsymbol{x}$, and a parameter $\boldsymbol{\theta}$. Since (1) is an approximation the interpretation of $\boldsymbol{\theta}$ is unclear; we *define* this target parameter by

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

where $\mu(d\boldsymbol{x})$ represents either Lebesgue measure or counting measure, depending upon the nature of the *design space* $\mathcal{X}$ with $\int_{\mathcal{X}} \mu(d\boldsymbol{x}) < \infty$. We then define $\psi(\boldsymbol{x}) = E\left[Y(\boldsymbol{x})\right] - \boldsymbol{f}'(\boldsymbol{x})\boldsymbol{\theta}$. This results in the class of responses $E\left[Y(\boldsymbol{x})\right] = \boldsymbol{f}'(\boldsymbol{x})\boldsymbol{\theta} + \psi(\boldsymbol{x})$, with – by virtue of (2) – ψ satisfying the orthogonality requirement

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

Assuming that $\mathcal{X}$ is rich enough that the matrix $\boldsymbol{A} = \int_{\mathcal{X}} \boldsymbol{f}(\boldsymbol{x})\boldsymbol{f}'(\boldsymbol{x})\mu(d\boldsymbol{x})$ is invertible, the parameter defined by (2) and (3) is unique.

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

$$\boldsymbol{M}_\xi = \int_{\mathcal{X}} \boldsymbol{f}(\boldsymbol{x})\boldsymbol{f}'(\boldsymbol{x})\xi(d\boldsymbol{x}), \quad \boldsymbol{b}_{\psi,\xi} = \int_{\mathcal{X}} \boldsymbol{f}(\boldsymbol{x})\psi(\boldsymbol{x})\xi(d\boldsymbol{x}),$$

and assume ξ is such that $\boldsymbol{M}_\xi$ is invertible. For an anticipated design of n, not necessarily distinct, points the covariance matrix of the least squares estimator $\hat{\boldsymbol{\theta}}$, assuming homoscedastic errors with variance σ_ε^2, is $\left(\sigma_\varepsilon^2/n\right)\boldsymbol{M}_\xi^{-1}$, and the bias is $E\left[\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\right] = \boldsymbol{M}_\xi^{-1}\boldsymbol{b}_{\psi,\xi}$;

together these yield the mean squared error (*mse*) matrix

$$\text{MSE}\left[\hat{\boldsymbol{\theta}}\right] = \frac{\sigma_\varepsilon^2}{n}\boldsymbol{M}_\xi^{-1} + \boldsymbol{M}_\xi^{-1}\boldsymbol{b}_{\psi,\xi}\boldsymbol{b}'_{\psi,\xi}\boldsymbol{M}_\xi^{-1}$$

of the parameter estimates, whence the *mse* of the predicted values $\hat{Y}(\boldsymbol{x}) = \boldsymbol{f}'(\boldsymbol{x})\hat{\boldsymbol{\theta}}$ is

$$\text{MSE}\left[\hat{Y}(\boldsymbol{x})\right] = \frac{\sigma_\varepsilon^2}{n}\boldsymbol{f}'(\boldsymbol{x})\boldsymbol{M}_\xi^{-1}\boldsymbol{f}(\boldsymbol{x}) + \left(\boldsymbol{f}'(\boldsymbol{x})\boldsymbol{M}_\xi^{-1}\boldsymbol{b}_{\psi,\xi}\right)^2 + \psi^2(\boldsymbol{x}).$$

A loss function that is commonly employed is the *integrated mse* of the *predictions*:

$$\begin{aligned}\text{IMSE}(\xi|\psi) &= \int_\chi \text{MSE}\left[\hat{Y}(\boldsymbol{x})\right]\mu(d\boldsymbol{x}) \\ &= \frac{\sigma_\varepsilon^2}{n}tr\left(\boldsymbol{A}\boldsymbol{M}_\xi^{-1}\right) + \boldsymbol{b}'_{\psi,\xi}\boldsymbol{M}_\xi^{-1}\boldsymbol{A}\boldsymbol{M}_\xi^{-1}\boldsymbol{b}_{\psi,\xi} + \int_\chi \psi^2(\boldsymbol{x})\mu(d\boldsymbol{x}). \end{aligned} \tag{4}$$

The dependence on ψ is eliminated by adopting a *minimax* approach, according to which one first maximizes (4) over a neighbourhood of the assumed response, i.e. over ψ. This neighbourhood is constrained by (3) and by a bound $\int_\chi \psi^2(\boldsymbol{x})\mu(d\boldsymbol{x}) \le \tau^2/n$, required so that errors due to bias and to variation remain of the same order, asymptotically. Define $\psi_0(\boldsymbol{x}) = \sqrt{n}\psi(\boldsymbol{x})/\tau$ and $\nu = \tau^2/\left(\sigma_\varepsilon^2 + \tau^2\right)$. Then $\max_\psi \text{IMSE}(\xi|\psi)$ is $\left(\sigma_\varepsilon^2 + \tau^2\right)/n$ times

$$I_\nu(\xi) = (1-\nu)\,\text{VAR}(\xi) + \nu\,\text{MAXBIAS}(\xi),$$

where $\text{VAR}(\xi) = tr\boldsymbol{A}\boldsymbol{M}_\xi^{-1}$ is the integrated variance of the predictors and $\text{MAXBIAS}(\xi) = \max_{\psi_0}\text{BIAS}(\xi|\psi_0)$, where

$$\text{BIAS}(\xi|\psi_0) = \boldsymbol{b}'_{\psi_0,\xi}\boldsymbol{M}_\xi^{-1}\boldsymbol{A}\boldsymbol{M}_\xi^{-1}\boldsymbol{b}_{\psi_0,\xi} + 1 \tag{5}$$

is the integrated (squared) bias, with ψ_0 constrained by (3) and by $\int_\chi \psi_0^2(\boldsymbol{x})\mu(d\boldsymbol{x}) = 1$.

3. Robust Bounded Maximum Bias and Bounded Variance designs

Let Ξ be a class of designs on χ, for instance all probability measures on $[-1,1]$ – requiring appropriate approximations to make them implementable – or all exact designs, i.e. with integer allocations, on a finite design space. For

$$\min_{\xi\in\Xi}\text{MAXBIAS}(\xi) \le b^2 \le \max_{\xi\in\Xi}\text{MAXBIAS}(\xi), \tag{6}$$

consider the problem

(B): Minimize $\text{VAR}(\xi)$ in the class $\Xi_b \subset \Xi$ of designs for which $\text{MAXBIAS}(\xi) \le b^2$.

We call a solution to (B) a *Robust Bounded Bias* design with bias bound b^2, denoted $\text{RBB}(b^2)$. For

$$\min_{\xi\in\Xi}\text{VAR}(\xi) \le s^2 \le \max_{\xi\in\Xi}\text{VAR}(\xi), \tag{7}$$

consider the problem

(S): Minimize $\text{MAXBIAS}(\xi)$ in the class $\Xi_s \subset \Xi$ of designs for which $\text{VAR}(\xi) \leq s^2$.

We call a solution to (S) a *Robust Bounded Variance* design with variance bound s^2, denoted $\text{RBV}(s^2)$.

For $\nu \in [0, 1]$ define $\xi_\nu = \arg\min_\xi I_\nu(\xi)$. Set $b^2(\nu) = \text{MAXBIAS}(\xi_\nu)$ and $s^2(\nu) = \text{VAR}(\xi_\nu)$. Note that then

$$\xi_\nu \in \Xi_{b(\nu)} \cap \Xi_{s(\nu)}.$$

In §3.1 we prove and discuss Theorem 1. This asserts that solutions to (B) and (S) are given by ξ_ν, for some $\nu \in [0, 1]$, for any b^2 and any s^2. Conversely, any ξ_ν is a solution to (B) and (S) for appropriate values of b^2 and s^2.

Theorem 1. *(a) Robust Bounded Bias designs with bias bound b^2 satisfying (6) are given by*

$$\xi = \begin{cases} \xi_\nu, & \text{if } b^2 = b^2(\nu), \text{ for } 0 \leq \nu \leq 1; \\ \xi_0, & \text{if } b^2 \geq b^2(0). \end{cases}$$

(b) Robust Bounded Variance designs with variance bound s^2 satisfying (7) are given by

$$\xi = \begin{cases} \xi_\nu, & \text{if } s^2 = s^2(\nu), \text{ for } 0 \leq \nu \leq 1; \\ \xi_1, & \text{if } s^2 \geq s^2(1). \end{cases}$$

3.1. Proof of Theorem 1

By the *I-optimal* design Studden (1977) we mean the minimizer of $I_0(\xi)$, i.e. of the Integrated Variance of the Predicted Values. By the *uniform* design we mean the design $\xi(d\boldsymbol{x}) \propto \mu(d\boldsymbol{x})$. These designs play special roles – they turn out to be ξ_0 and ξ_1, respectively.

Lemma 1. *The design ξ_0, minimizing $I_0(\xi) = \text{VAR}(\xi)$ in Ξ, is I-optimal and the design ξ_1, minimizing $I_1(\xi) = \text{MAXBIAS}(\xi)$ in Ξ is uniform.*

Proof: That ξ_0 is the I-optimal design follows from the definition: $I_0(\xi) = \text{VAR}(\xi)$. By (5), $\text{MAXBIAS}(\xi) \geq 1$. This lower bound is attained by the uniform design, since then

$$\boldsymbol{b}_{\psi,\xi} = \int_{\mathcal{X}} \boldsymbol{f}(\boldsymbol{x})\,\psi(\boldsymbol{x})\,\xi(d\boldsymbol{x}) \propto \int_{\mathcal{X}} \boldsymbol{f}(\boldsymbol{x})\,\psi(\boldsymbol{x})\,\mu(d\boldsymbol{x}) = \boldsymbol{0},$$

by (3). □

By Lemma 1, the lower bounds of the ranges (6) and (7) are attained by ξ_1 and ξ_0 respectively.

Lemma 2. *(a) ξ_ν is a solution to (B) for $b^2(\nu)$. (b) If $b^2 \geq b^2(0) = \text{MAXBIAS}(\xi_0)$ then ξ_0 is a solution to (B) for b^2.*

Proof: (a) For any $\xi \in \Xi_{b(\nu)}$ we must have that $\text{VAR}(\xi) \geq \text{VAR}(\xi_\nu)$, since otherwise

$$I_\nu(\xi) = (1-\nu)\,\text{VAR}(\xi) + \nu\text{MAXBIAS}(\xi) < (1-\nu)\,\text{VAR}(\xi_\nu) + \nu b^2(\nu) = I_\nu(\xi_\nu),$$

a contradiction. Thus $\xi_\nu \in \Xi_{b(\nu)}$ minimizes $\text{VAR}(\xi)$ in $\Xi_{b(\nu)}$, i.e. is $\text{RBB}(b^2(\nu))$.
(b) For such b^2 we have that $\text{MAXBIAS}(\xi_0) \leq b^2$ so that $\xi_0 \in \Xi_b$. As well,

$$\text{VAR}(\xi_0) = \min_{\xi\in\Xi} \text{VAR}(\xi) \leq \min_{\xi\in\Xi_b} \text{VAR}(\xi) \leq \text{VAR}(\xi_0), \tag{8}$$

so that we must have equality throughout in (8) and ξ_0 is $\text{RBB}(b^2)$. □

Lemma 3. *(a) ξ_ν is a solution to (S) for $s^2(\nu)$. (b) If $s^2 \geq s^2(1) = \text{VAR}(\xi_1)$ then ξ_1 is a solution to (S) for s^2.*

The proof of Lemma 3 is identical to that of Lemma 2, apart from the obvious interchanges $\text{MAXBIAS} \leftrightarrow \text{VAR}$, $b \leftrightarrow s$, $\xi_0 \leftrightarrow \xi_1$, and so is omitted. Lemmas 1, 2 and 3 together prove Theorem 1.

Remark: The solutions in Theorem 1 suggest that the maxima of $b^2(\cdot)$ and $s^2(\cdot)$ are unique, with

$$\text{(i) } \{0\} = \arg\max_{\nu\in[0,1]} b^2(\nu) \text{ and (ii) } \{1\} = \arg\max_{\nu\in[0,1]} s^2(\nu). \tag{9}$$

If not, for instance if $b^2(\nu)$ has multiple maxima or if '0' is not a maximum, then on the set $N_b = \left\{\nu^* \in [0,1] \,\middle|\, b^2(\nu^*) \geq b^2(0)\right\}$, both ξ_{ν^*} and ξ_0 are $\text{RBB}\left(b^2(\nu^*)\right)$, by (a) of Theorem 1 and Lemma 1 respectively, hence furnish the same minimum variance. Similarly, if (ii) fails then the bias is constant on the corresponding set N_s, where $s^2(\nu^*) \geq s^2(1)$. While counterintuitive if these sets are not singletons – and if the design weights ξ_i vary continuously with ν – these events cannot be ruled out without further restrictions. This is shown by the example of regression through the origin – $p = q = 1$, $f(x) = x$, and Ξ the class of designs placing mass $\{\alpha, 1-2\alpha, \alpha\}$ on the points of $\chi = \{-1, 0, 1\}$. For any such design we find that $\text{VAR}(\xi) = 1/\alpha$ and $\text{MAXBIAS}(\xi) = 1$, so that $I_\nu(\xi) = (1-\nu)/\alpha + \nu$ is minimized by $\xi_\nu = .5\delta_{\pm 1}$ for any ν. Thus $I_\nu(\xi_\nu) = 2 - \nu$, (9) fails and in fact $b^2(\cdot)$ and $s^2(\cdot)$ are constant on $[0,1]$: $b^2(\nu) = \text{MAXBIAS}(\xi_\nu) \equiv 1$, $s^2(\nu) = \text{VAR}(\xi_\nu) \equiv 2$.

4. Examples

We now assume that the design space is finite: $\mathcal{X} = \{\boldsymbol{x}_1, \cdots, \boldsymbol{x}_N\}$. Let $\boldsymbol{Q}$ be an $N \times p$ matrix whose orthonormal columns span the column space of $(\boldsymbol{f}(\boldsymbol{x}_1), \cdots, \boldsymbol{f}(\boldsymbol{x}_N))'$, assumed to have dimension p. For a design ξ placing mass ξ_i on $\boldsymbol{x}_i$ define $\boldsymbol{D}(\xi) = diag(\xi_1, \cdots, \xi_N)$. Then in terms of

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

it is shown in Wiens (2018) that

$$\text{VAR}(\xi) = tr\boldsymbol{R}^{-1}(\xi),\ \text{MAXBIAS}(\xi) = ch_{\max}\boldsymbol{U}(\xi).$$

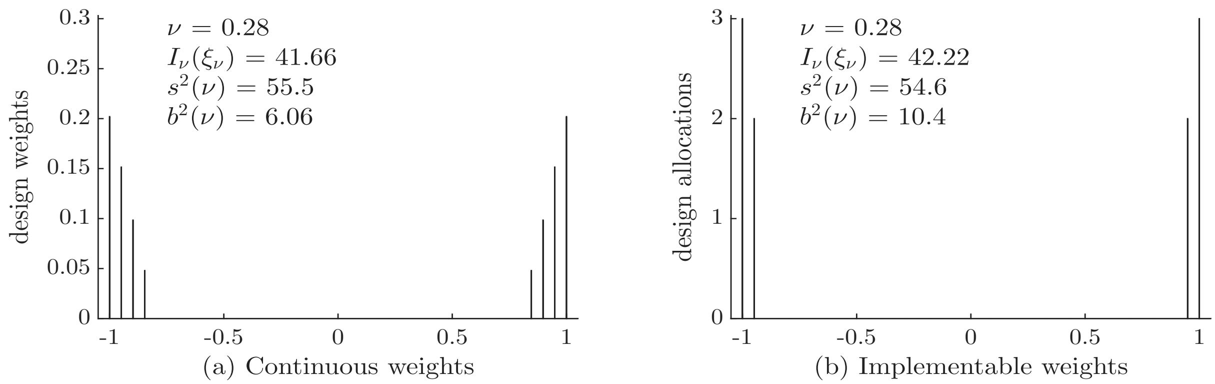

Figure 1: (a) RBB(b^2(.28))/RBV(s^2(.28)) designs for the values displayed (CMB = .33). (b) Implementation of the design in (a); design size n = 10.

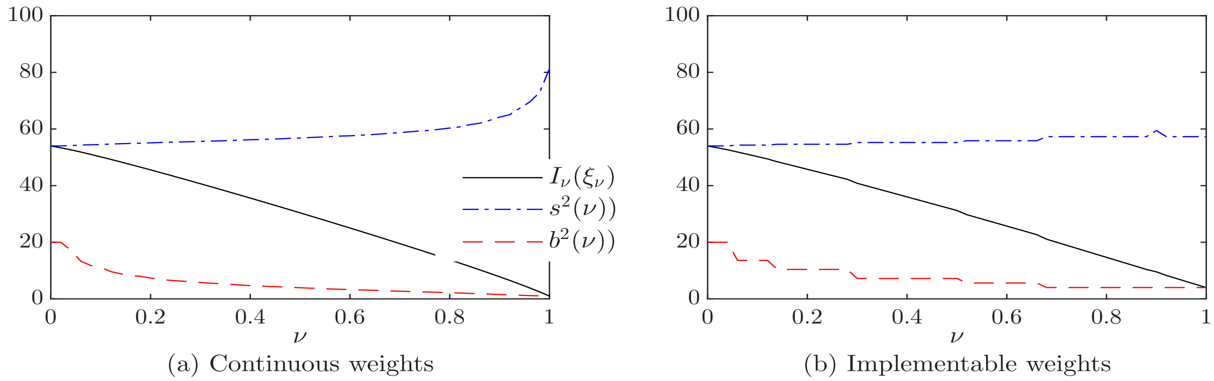

Figure 2: IMSE, VAR and MAXBIAS vs. ν for the continuous optimal designs and their implementable approximations (n = 10).

Here *tr* and $ch_{\max}$ denote the trace and maximum eigenvalue, respectively.

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

$$I_\nu\left(\xi_{n+1}^{(i)}\right) = I_\nu(\xi_n) - t_{n,i}/n + O\left(n^{-2}\right), \tag{10}$$

and then $\boldsymbol{x}_{(i)}$, with $(i) = \arg\max_i t_{n,i}$, is added to the design. This is carried out to convergence.

A useful measure in choosing a value of the bias/variance parameter ν is the dimensionless *coefficient of maximum bias*, A useful measure in choosing a value of the bias/variance parameter ν is the dimensionless *coefficient of maximum bias*, defined by

$$\text{CMB}(\nu) = \sqrt{b^2(\nu)/s^2(\nu)}.$$

We first present output for the case of straight line regression on a symmetric design space of size $N = 40$ in $[-1, 1]$: $\chi = \{x_i \,|\, i = 1, ..., N\}$ with $x_i = -1 + 2(i-1)/(N-1) =$

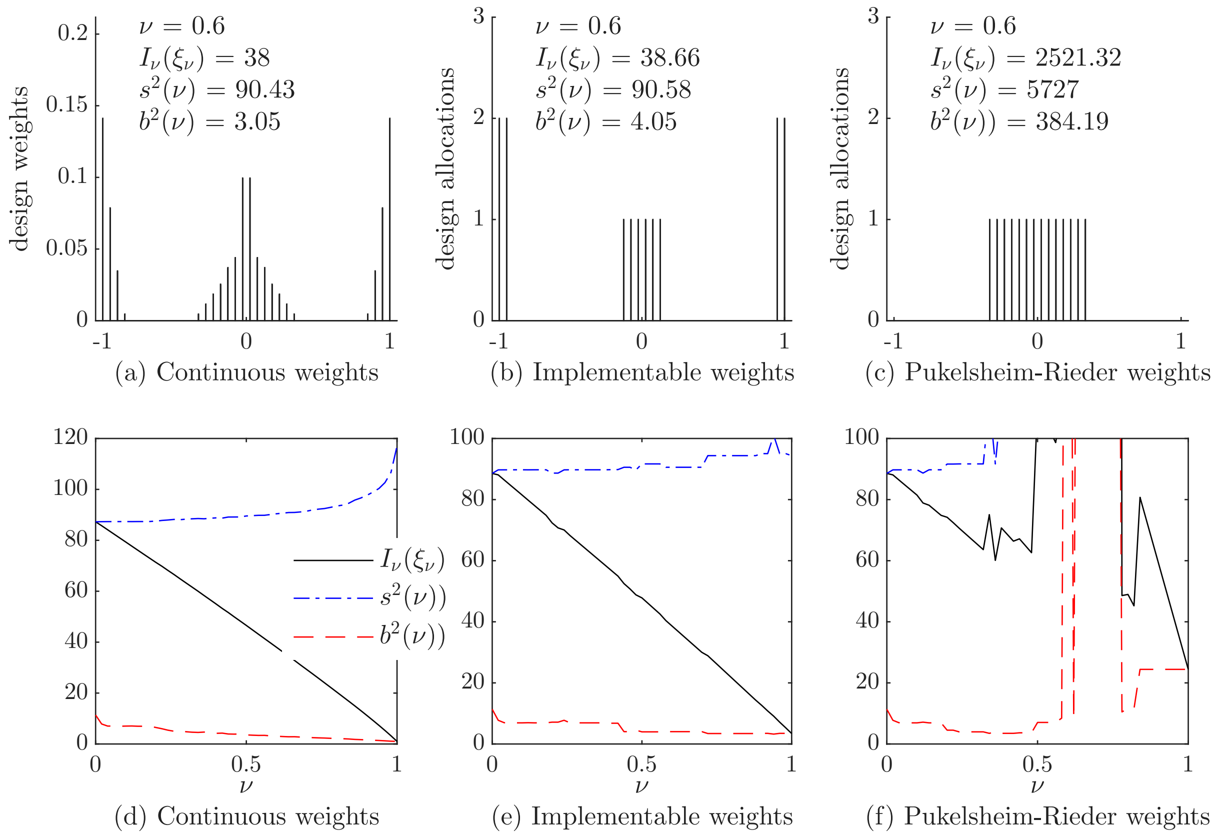

Figure 3: (a) – (c): Continuous and implementable designs for quadratic regression; $n = 14$. (d) – (f) IMSE, VAR and MAXBIAS VS. ν

$-x_{N-i+1}$. Put $\boldsymbol{x} = (x_1, ..., x_N)'$. Then $\boldsymbol{Q}_{N\times 2} = \left(\mathbf{1}_N / \sqrt{N} \vdots \boldsymbol{x} / \sqrt{\|\boldsymbol{x}\|}\right)$. To obtain the RBB and RBV designs we minimize $I_\nu(\xi)$ over unrestricted designs ξ, so that the ξ_i vary freely over the simplex $\mathcal{S} = \{\xi_i \in [0,1], \sum \xi_i = 1\}$. In particular, the minimax design when $\nu = 1$ is uniform: $\xi_{1,i} \equiv 1/N$, so that $\boldsymbol{U}(\xi_1) = \boldsymbol{I}_N$ with BIAS$(\xi_1) = 1$ and VAR$(\xi_1) = 2N$. When $\nu = 0$, the minimax design is $\xi_0 = .5\delta_{\pm 1}$, with

$$\text{MAXBIAS}(\xi_0) = \max\left(N/2, \|\boldsymbol{x}\|^2/2\right) = N/2, \quad \text{VAR}(\xi_0) = N + \|\boldsymbol{x}\|^2.$$

Thus ξ_1 is RBB$(b^2 = 1)$ and ξ_0 is RBV$(s^2 = N + \|\boldsymbol{x}\|^2)$. Representative results are presented in parts (a) of Figure 1 (for which we specified CMB$(\nu) \cong 1/3$ and obtained $\nu = .28$) and Figure 2 – note that (9) holds.

The design in (a) of Figure 1 is not implementable since the allocations $n_i = n\xi_i$ need not be integers. To obtain the implementation in part (b) of this figure we first rounded up the n_i to $\lceil n\xi_i \rceil$, whose sum then exceeds n. The excess is decreased stepwise, by removing points whose value of $t_{n,i}$ in (10) is a minimum. This method typically results in only a very small increase in the minimized value of the IMSE.

The behaviour shown in (b) of Figure 2, in particular of the bias and as anticipated in the Remark of §3.1, reflects the lack of continuity of the allocations as functions of ν.

Together with CMB(·) this plot can also serve as a guide to the designer in choosing a value of ν for an implementable design.

Remark: Our method of rounding the design weights so as to obtain implementable designs is somewhat non-standard, and is intended to preserve, as much as possible, the minimized IMSE. A more common method is the 'efficient design apportionment' method of Pukelsheim and Rieder (1992). This is a rounding procedure that has, amongst others, the property of 'sample size monotonicity' – if a new point is to be added to an existing design, then none of the current allocations will be reduced. Unless this property is required we cannot recommend this method in the current application, as it is too often very unstable, resulting in large increases in the loss. An example is quadratic regression on the same design space as above, with a design size $n = 14$, as illustrated in Figure 3.

Acknowledgements

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

References

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.

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

Pukelsheim, F. and Rieder, S. (1992), "Efficient Rounding of Approximate Designs," *Biometrika*, 79, 763–770.

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. (1990), "Robust, Minimax Designs for Multiple Linear Regression," *Linear Algebra and Its Applications, Second Special Issue on Linear Algebra and Statistics*, 127, 327–340.

— (1992), "Minimax Designs for Approximately Linear Regression," *Journal of Statistical Planning and Inference*, 31, 353–371.

— (2015), "Robustness of Design," in *Handbook of Design and Analysis of Experiments*, eds. Dean, A., Morris, M., Stufken, J., and Bingham, D., Chapman & Hall/CRC, pp. 287–303.

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