Notes on Robust Bayesian Design

Douglas P. Wiens[1]

July 2, 2019

1 Introduction

Suppose that, given a standard regression model $\mathbf{Y}_{n\times 1} = \mathbf{X}_{n\times p}\boldsymbol{\theta}$ + error, in which the rows of the model matrix $\mathbf{X}$ are of the form $\{\mathbf{f}'(\mathbf{x}_j)\}_{j=1}^n$ for p-dimensional regressors $\mathbf{f}(\mathbf{x})$, we make the moment assumptions

$$E\left[\mathbf{Y}|\boldsymbol{\theta},\sigma^2\right] = \mathbf{X}\boldsymbol{\theta},\ \text{COV}\left[\mathbf{Y}|\sigma^2,\boldsymbol{\theta}\right] = \sigma^2\mathbf{I}_n,$$

with prior moments

$$E\left[\boldsymbol{\theta}|\sigma^2\right] = \boldsymbol{\theta}_0,\ \text{COV}\left[\boldsymbol{\theta}|\sigma^2\right] = \frac{\sigma^2}{n_0}\mathbf{C}_0,\ E\left[\sigma^2\right] = \sigma_0^2.$$

Assume that $\mathbf{C}_0 > \mathbf{0}$.

In this framework Pukelsheim (1993, p. 268 ff.) derives the minimum mean square error affine predictor of $\mathbf{f}'(\mathbf{x})\boldsymbol{\theta}$. This predictor is of the form $\boldsymbol{\alpha}'(\mathbf{x})\mathbf{Y} + \beta(\mathbf{x})$ and minimizes

$$\text{MSE}\left(\boldsymbol{\alpha}(\mathbf{x}),\beta(\mathbf{x})\right) \stackrel{def}{=} E\left[\left(\boldsymbol{\alpha}'(\mathbf{x})\mathbf{Y} + \beta(\mathbf{x}) - \mathbf{f}'(\mathbf{x})\boldsymbol{\theta}\right)^2\right].$$

The calculations – see Appendix – give the minimizers

$$\boldsymbol{\alpha}_0(\mathbf{x}) = \mathbf{X}\left(n_0\mathbf{C}_0^{-1} + \mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{f}(\mathbf{x}), \tag{1a}$$

$$\beta_0(\mathbf{x}) = \mathbf{f}'(\mathbf{x})\left(n_0\mathbf{C}_0^{-1} + \mathbf{X}'\mathbf{X}\right)^{-1} n_0\mathbf{C}_0^{-1}\boldsymbol{\theta}_0, \tag{1b}$$

and the minimum MSE is

$$\text{MSE}_0\left(\boldsymbol{\alpha}_0(\mathbf{x}),\beta_0(\mathbf{x})\right) = \sigma_0^2\mathbf{f}'(\mathbf{x})\left(n_0\mathbf{C}_0^{-1} + \mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{f}(\mathbf{x}). \tag{2}$$

Now consider the alternative $E\left[Y(\mathbf{x})|\boldsymbol{\theta},\sigma^2\right] = \mathbf{f}'(\mathbf{x})\boldsymbol{\theta} + \psi(\mathbf{x})$. We suppose that the experimenter will act as though $\psi \equiv 0$, hence predict using $\boldsymbol{\alpha}_0$ and β_0, but wishes protection, through the design, against increased losses arising from model misspecification. The design space $\chi = \{\mathbf{x}_i\}_{i=1}^N$ is finite.

In order that the alternate models be well-defined we impose the condition

$$\sum_{i=1}^N \mathbf{f}(\mathbf{x}_i)\psi(\mathbf{x}_i) = \mathbf{0}; \tag{3}$$

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

as well we assume that

$$\sum_{i=1}^{N} \psi^2(\mathbf{x}_i) \leq \tau^2/n, \tag{4}$$

for a finite constant τ. We use the notation $\mathbf{F}_{N\times p} = (\mathbf{f}(\mathbf{x}_1), \cdots, \mathbf{f}(\mathbf{x}_N))'$ and $\boldsymbol{\psi}_{N\times 1} = (\psi(\mathbf{x}_1), \cdots, \psi(\mathbf{x}_N))'$, so that (3) and (4) become, respectively,

$$\text{(i) } \mathbf{F}'\boldsymbol{\psi} = \mathbf{0}, \text{ (ii) } \|\boldsymbol{\psi}\|^2 \leq \tau^2/n. \tag{5}$$

For a design $\boldsymbol{\xi}$ on χ, placing mass n_i/n at $\mathbf{x}_i$ ($\sum_{i=1}^{N} n_i = n$), we define

$$\begin{aligned}
\mathbf{D}(\boldsymbol{\xi}) &= diag(\xi_1, \cdots, \xi_N), \\
\mathbf{M}_{p\times p} &= \sum_{i=1}^{N} \xi_i \mathbf{f}(\mathbf{x}_i)\mathbf{f}'(\mathbf{x}_i) = \mathbf{F}'\mathbf{D}(\boldsymbol{\xi})\mathbf{F}, \\
\mathbf{b}_{p\times 1} &= \sum_{i=1}^{N} \xi_i \mathbf{f}(\mathbf{x}_i)\psi(\mathbf{x}_i) = \mathbf{F}'\mathbf{D}(\boldsymbol{\xi})\boldsymbol{\psi}.
\end{aligned} \tag{6}$$

Note that $\mathbf{X}'\mathbf{X} = n\mathbf{M}$.

The MSE becomes

$$\text{MSE}(\mathbf{x};\psi) = E\left[\left(\boldsymbol{\alpha}_0'(\mathbf{x})\mathbf{Y} + \beta_0(\mathbf{x}) - (\mathbf{f}'(\mathbf{x})\boldsymbol{\theta} + \psi(\mathbf{x}))\right)^2\right].$$

After a calculation (detailed in the Appendix), we find that

$$\text{MSE}(\mathbf{x};\psi) = \sigma_0^2 \mathbf{f}'(\mathbf{x})\left(n_0\mathbf{C}_0^{-1} + n\mathbf{M}\right)^{-1}\mathbf{f}(\mathbf{x}) + \left(\mathbf{f}'(\mathbf{x})\left(n_0\mathbf{C}_0^{-1} + n\mathbf{M}\right)^{-1} n\mathbf{b} - \psi(\mathbf{x})\right)^2.$$

Our loss function is

$$\begin{aligned}
&\text{IMSE}(\psi) \\
&= \sum_{i=1}^{N} \text{MSE}(\mathbf{x}_i;\psi) \\
&= \sigma_0^2 tr\left(n_0\mathbf{C}_0^{-1} + n\mathbf{M}\right)^{-1}\mathbf{F}'\mathbf{F} + n^2\mathbf{b}'\left(n_0\mathbf{C}_0^{-1} + n\mathbf{M}\right)^{-1}\mathbf{F}'\mathbf{F}\left(n_0\mathbf{C}_0^{-1} + n\mathbf{M}\right)^{-1}\mathbf{b} + \|\boldsymbol{\psi}\|^2.
\end{aligned}$$

It is convenient to work in a canonical form. Let $\mathbf{Q}_{N\times p}$ be such that its columns form an orthogonal basis for the column space of $\mathbf{F}$ – this is obtained via the QR-decomposition $\mathbf{F} = \mathbf{QR}$ for a nonsingular, triangular $\mathbf{R}$. We will require as well the $N \times N - p$ matrix $\mathbf{Q}_+$ extending $\mathbf{Q}$ to a square, orthogonal matrix $\left(\mathbf{Q} \vdots \mathbf{Q}_+\right) : N \times N$ (so that $\mathbf{Q}_+\mathbf{Q}_+' = \mathbf{I}_n - \mathbf{QQ}'$). If we now define a $p \times p$ matrix $\mathbf{M}_0$ by $\mathbf{C}_0^{-1} = \mathbf{R}'\mathbf{M}_0\mathbf{R}$, and define as well

$$\begin{aligned}
\mathbf{V}(\boldsymbol{\xi}) &= \mathbf{Q}'\mathbf{D}(\boldsymbol{\xi})\mathbf{Q}, \\
\mathbf{W}(\boldsymbol{\xi}) &= \mathbf{Q}'\mathbf{D}^2(\boldsymbol{\xi})\mathbf{Q},
\end{aligned}$$

then

$$n_0\mathbf{C}_0^{-1} + n\mathbf{M} = \mathbf{R}'(n_0\mathbf{M}_0 + n\mathbf{V}(\boldsymbol{\xi}))\mathbf{R},$$

and we obtain

$$\begin{aligned}\text{IMSE}\left(\psi\right) &= \sigma_0^2 tr\left(n_0\mathbf{M}_0 + n\mathbf{V}\left(\boldsymbol{\xi}\right)\right)^{-1} \\ &\quad + n^2\boldsymbol{\psi}'\mathbf{D}\left(\boldsymbol{\xi}\right)\mathbf{Q}\left(n_0\mathbf{M}_0 + n\mathbf{V}\left(\boldsymbol{\xi}\right)\right)^{-2}\mathbf{Q}'\mathbf{D}\left(\boldsymbol{\xi}\right)\boldsymbol{\psi} + \left\|\boldsymbol{\psi}\right\|^2.\end{aligned}$$

By virtue of (5), $\boldsymbol{\psi} = \mathbf{Q}_+\mathbf{c}$ for some $\mathbf{c}_{N-p\times 1}$, where $\|\mathbf{c}\| = \|\boldsymbol{\psi}\| = \tau/\sqrt{n}$. Thus

$$\begin{aligned}&\text{IMSE}\left(\psi\right) \\ &= \frac{\sigma_0^2}{n} tr\left(\frac{n_0}{n}\mathbf{M}_0 + \mathbf{V}\left(\boldsymbol{\xi}\right)\right)^{-1} + \mathbf{c}'\mathbf{Q}_+'\mathbf{D}\left(\boldsymbol{\xi}\right)\mathbf{Q}\left(\frac{n_0}{n}\mathbf{M}_0 + \mathbf{V}\left(\boldsymbol{\xi}\right)\right)^{-2}\mathbf{Q}'\mathbf{D}\left(\boldsymbol{\xi}\right)\mathbf{Q}_+\mathbf{c} + \left\|\mathbf{c}\right\|^2,\end{aligned}$$

with

$$\begin{aligned}&\max_{\psi}\text{IMSE}\left(\psi\right) \\ &= \frac{\sigma_0^2}{n} tr\left(\frac{n_0}{n}\mathbf{M}_0 + \mathbf{V}\left(\boldsymbol{\xi}\right)\right)^{-1} + \frac{\tau^2}{n} ch_{\max}\mathbf{Q}_+'\mathbf{D}\left(\boldsymbol{\xi}\right)\mathbf{Q}\left(\frac{n_0}{n}\mathbf{M}_0 + \mathbf{V}\left(\boldsymbol{\xi}\right)\right)^{-2}\mathbf{Q}'\mathbf{D}\left(\boldsymbol{\xi}\right)\mathbf{Q}_+ + \frac{\tau^2}{n} \\ &= \frac{\sigma_0^2}{n} tr\left(\frac{n_0}{n}\mathbf{M}_0 + \mathbf{V}\left(\boldsymbol{\xi}\right)\right)^{-1} \\ &\quad + \frac{\tau^2}{n} ch_{\max}\left\{\left(\frac{n_0}{n}\mathbf{M}_0 + \mathbf{V}\left(\boldsymbol{\xi}\right)\right)^{-1}\mathbf{Q}'\mathbf{D}\left(\boldsymbol{\xi}\right)\mathbf{Q}_+\mathbf{Q}_+'\mathbf{D}\left(\boldsymbol{\xi}\right)\mathbf{Q}\left(\frac{n_0}{n}\mathbf{M}_0 + \mathbf{V}\left(\boldsymbol{\xi}\right)\right)^{-1}\right\} + \frac{\tau^2}{n} \\ &= \frac{\sigma_0^2}{n} tr\left(\frac{n_0}{n}\mathbf{M}_0 + \mathbf{V}\left(\boldsymbol{\xi}\right)\right)^{-1} \\ &\quad + \frac{\tau^2}{n}\left[1 + ch_{\max}\left\{\left(\frac{n_0}{n}\mathbf{M}_0 + \mathbf{V}\left(\boldsymbol{\xi}\right)\right)^{-1}\left[\mathbf{W}\left(\boldsymbol{\xi}\right) - \mathbf{V}^2\left(\boldsymbol{\xi}\right)\right]\left(\frac{n_0}{n}\mathbf{M}_0 + \mathbf{V}\left(\boldsymbol{\xi}\right)\right)^{-1}\right\}\right].\end{aligned}$$

Thus $\max_{\psi}\text{IMSE}(\psi) = \frac{\sigma_0^2+\tau^2}{n}$ times

$$\begin{aligned}\mathcal{L}_\nu\left(\xi\right) &= \left(1-\nu\right) tr\left(\frac{n_0}{n}\mathbf{M}_0 + \mathbf{V}\left(\boldsymbol{\xi}\right)\right)^{-1} \\ &\quad + \nu\left(1 + ch_{\max}\left\{\left(\frac{n_0}{n}\mathbf{M}_0 + \mathbf{V}\left(\boldsymbol{\xi}\right)\right)^{-1}\left[\mathbf{W}\left(\boldsymbol{\xi}\right) - \mathbf{V}^2\left(\boldsymbol{\xi}\right)\right]\left(\frac{n_0}{n}\mathbf{M}_0 + \mathbf{V}\left(\boldsymbol{\xi}\right)\right)^{-1}\right\}\right),\end{aligned}$$

where $\nu = \tau^2/\left(\sigma_0^2 + \tau^2\right) \in [0,1]$.

Notes:

1. Maximize over a neighbourhood of $\mathbf{M}_0$, i.e., as neighbourhood of the prior covariance matrix of the regression parameters, as well?

2. As in Wiens (2018), if $\mathbf{V}\left(\boldsymbol{\xi}\right)$ is held fixed then the loss is convex in ξ and the parametric form of the minimizing design, subject to the constraints, can be found analytically. A numerical minimization of the loss over these parameters then yields the final, minimax design.

Appendix: Derivations

Details for (1) and (2): We are to minimize

$$mse\left(\boldsymbol{\alpha}(\mathbf{x}),\beta(\mathbf{x})\right)=E\left[Z^2(\mathbf{x})\right], \text{for } Z(\mathbf{x})=\boldsymbol{\alpha}'(\mathbf{x})\,\mathbf{Y}+\beta(\mathbf{x})-\mathbf{f}'(\mathbf{x})\,\boldsymbol{\theta}.$$

Conditional on $\boldsymbol{\theta},\sigma^2$,

$$\begin{aligned}E\left[Z^2(\mathbf{x})|\boldsymbol{\theta},\sigma^2\right] &= \text{VAR}\left[Z(\mathbf{x})|\boldsymbol{\theta},\sigma^2\right]+\left(E\left[Z(\mathbf{x})|\boldsymbol{\theta},\sigma^2\right]\right)^2\\ &= \sigma^2\boldsymbol{\alpha}'(\mathbf{x})\boldsymbol{\alpha}(\mathbf{x})+\left(\boldsymbol{\gamma}'(\mathbf{x})\boldsymbol{\theta}+\beta(\mathbf{x})\right)^2, \text{ for } \boldsymbol{\gamma}'(\mathbf{x})=\boldsymbol{\alpha}'(\mathbf{x})\,\mathbf{X}-\mathbf{f}'(\mathbf{x}).\end{aligned}$$

Averaging over $\boldsymbol{\theta}$:

$$\begin{aligned}E\left[\left(\boldsymbol{\gamma}'\boldsymbol{\theta}+\beta(\mathbf{x})\right)^2|\sigma^2\right] &= \text{VAR}\left[\left(\boldsymbol{\gamma}'(\mathbf{x})\boldsymbol{\theta}+\beta(\mathbf{x})\right)|\sigma^2\right]+\left(E\left[\left(\boldsymbol{\gamma}'(\mathbf{x})\boldsymbol{\theta}+\beta(\mathbf{x})\right)|\sigma^2\right]\right)^2\\ &= \frac{\sigma^2}{n_0}\boldsymbol{\gamma}'(\mathbf{x})\,\mathbf{C}_0\boldsymbol{\gamma}(\mathbf{x})+\left(\boldsymbol{\gamma}'(\mathbf{x})\boldsymbol{\theta}_0+\beta(\mathbf{x})\right)^2,\end{aligned}$$

and then averaging over σ^2 gives

$$mse\left(\boldsymbol{\alpha}(\mathbf{x}),\beta(\mathbf{x})\right)=\sigma_0^2\boldsymbol{\alpha}'(\mathbf{x})\boldsymbol{\alpha}(\mathbf{x})+\frac{\sigma_0^2}{n_0}\boldsymbol{\gamma}'(\mathbf{x})\,\mathbf{C}_0\boldsymbol{\gamma}(\mathbf{x})+\left(\boldsymbol{\gamma}'(\mathbf{x})\boldsymbol{\theta}_0+\beta(\mathbf{x})\right)^2.$$

Minimizing over $\beta(\mathbf{x})$ gives $\beta_0(\mathbf{x})=-\boldsymbol{\gamma}'(\mathbf{x})\boldsymbol{\theta}=\left(\mathbf{f}'(\mathbf{x})-\boldsymbol{\alpha}'(\mathbf{x})\,\mathbf{X}\right)\boldsymbol{\theta}_0$, and

$$\begin{aligned}mse\left(\boldsymbol{\alpha}(\mathbf{x}),\beta_0(\mathbf{x})\right) &= \sigma_0^2\boldsymbol{\alpha}'(\mathbf{x})\boldsymbol{\alpha}(\mathbf{x})+\frac{\sigma_0^2}{n_0}\boldsymbol{\gamma}'(\mathbf{x})\,\mathbf{C}_0\boldsymbol{\gamma}(\mathbf{x})\\ &= \sigma_0^2\left\{\begin{array}{c}\boldsymbol{\alpha}'(\mathbf{x})\left(\mathbf{I}_n+\frac{1}{n_0}\mathbf{X}\mathbf{C}_0\mathbf{X}'\right)\boldsymbol{\alpha}(\mathbf{x})-\frac{2}{n_0}\boldsymbol{\alpha}'(\mathbf{x})\,\mathbf{X}\mathbf{C}_0\mathbf{f}(\mathbf{x})\\ +\frac{1}{n_0}\mathbf{f}'(\mathbf{x})\,\mathbf{C}_0\mathbf{f}(\mathbf{x})\end{array}\right\}\\ &= \sigma_0^2\left\{\begin{array}{c}\left(\boldsymbol{\alpha}(\mathbf{x})-\left(\mathbf{I}_n+\frac{1}{n_0}\mathbf{X}\mathbf{C}_0\mathbf{X}'\right)^{-1}\frac{1}{n_0}\mathbf{X}\mathbf{C}_0\mathbf{f}(\mathbf{x})\right)'\\ \cdot\left(\mathbf{I}_n+\frac{1}{n_0}\mathbf{X}\mathbf{C}_0\mathbf{X}'\right)\cdot\\ \left(\boldsymbol{\alpha}(\mathbf{x})-\left(\mathbf{I}_n+\frac{1}{n_0}\mathbf{X}\mathbf{C}_0\mathbf{X}'\right)^{-1}\frac{1}{n_0}\mathbf{X}\mathbf{C}_0\mathbf{f}(\mathbf{x})\right)\end{array}\right\}\\ &\quad+\sigma_0^2\left\{\begin{array}{c}\frac{1}{n_0}\mathbf{f}'(\mathbf{x})\,\mathbf{C}_0\mathbf{f}(\mathbf{x})\\ -\left(\frac{1}{n_0}\mathbf{X}\mathbf{C}_0\mathbf{f}(\mathbf{x})\right)'\left(\mathbf{I}_n+\frac{1}{n_0}\mathbf{X}\mathbf{C}_0\mathbf{X}'\right)^{-1}\left(\frac{1}{n_0}\mathbf{X}\mathbf{C}_0\mathbf{f}(\mathbf{x})\right)\end{array}\right\}.\end{aligned}$$

A useful identity is

$$\left(\mathbf{I}_n+\frac{1}{n_0}\mathbf{X}\mathbf{C}_0\mathbf{X}'\right)^{-1}=\mathbf{I}_n-\mathbf{X}\left(n_0\mathbf{C}_0^{-1}+\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}', \tag{A.1}$$

implying

$$\left(\mathbf{I}_n+\frac{1}{n_0}\mathbf{X}\mathbf{C}_0\mathbf{X}'\right)^{-1}\frac{1}{n_0}\mathbf{X}\mathbf{C}_0=\mathbf{X}\left(n_0\mathbf{C}_0^{-1}+\mathbf{X}'\mathbf{X}\right)^{-1}. \tag{A.2}$$

Thus, with

$$\boldsymbol{\alpha}_0(\mathbf{x}) = \left(\mathbf{I}_n + \frac{1}{n_0}\mathbf{X}\mathbf{C}_0\mathbf{X}'\right)^{-1} \frac{1}{n_0}\mathbf{X}\mathbf{C}_0\mathbf{f}\left(\mathbf{x}\right) = \mathbf{X}\left(n_0\mathbf{C}_0^{-1} + \mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{f}\left(\mathbf{x}\right),$$

for which

$$\beta_0(\mathbf{x}) = \mathbf{f}'\left(\mathbf{x}\right)\left(n_0\mathbf{C}_0^{-1} + \mathbf{X}'\mathbf{X}\right)^{-1} n_0\mathbf{C}_0^{-1}\boldsymbol{\theta}_0,$$

we have that

$$\begin{aligned} mse\left(\boldsymbol{\alpha}(\mathbf{x}), \beta_0(\mathbf{x})\right) &= \sigma_0^2\left(\boldsymbol{\alpha}(\mathbf{x}) - \boldsymbol{\alpha}_0(\mathbf{x})\right)'\left(\mathbf{I}_n + \frac{1}{n_0}\mathbf{X}\mathbf{C}_0\mathbf{X}'\right)\left(\boldsymbol{\alpha}(\mathbf{x}) - \boldsymbol{\alpha}_0(\mathbf{x})\right) \\ &\quad + \sigma_0^2\left\{\frac{1}{n_0}\mathbf{f}'\left(\mathbf{x}\right)\mathbf{C}_0\mathbf{f}\left(\mathbf{x}\right) - \boldsymbol{\alpha}_0'(\mathbf{x})\left(\mathbf{I}_n + \frac{1}{n_0}\mathbf{X}\mathbf{C}_0\mathbf{X}'\right)\boldsymbol{\alpha}_0(\mathbf{x})\right\}. \end{aligned}$$

Thus the minimizers are $\boldsymbol{\alpha}_0(\mathbf{x})$ and $\beta_0(\mathbf{x})$, and the minimum MSE is

$$\begin{aligned} mse\left(\boldsymbol{\alpha}_0(\mathbf{x}), \beta_0(\mathbf{x})\right) &= \sigma_0^2\left\{\frac{1}{n_0}\mathbf{f}'\left(\mathbf{x}\right)\mathbf{C}_0\mathbf{f}\left(\mathbf{x}\right) - \boldsymbol{\alpha}_0'(\mathbf{x})\left(\mathbf{I}_n + \frac{1}{n_0}\mathbf{X}\mathbf{C}_0\mathbf{X}'\right)\boldsymbol{\alpha}_0(\mathbf{x})\right\} \\ &= \sigma_0^2\mathbf{f}'\left(\mathbf{x}\right)\left(n_0\mathbf{C}_0^{-1} + \mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{f}\left(\mathbf{x}\right). \end{aligned}$$

□

Derivation of MSE($\mathbf{x};\psi$): We have MSE$(\mathbf{x};\psi) = E\left[Z^2\left(\mathbf{x}\right)\right]$, for

$$Z\left(\mathbf{x}\right) = \boldsymbol{\alpha}_0'(\mathbf{x})\,\mathbf{Y} + \beta_0(\mathbf{x}) - \left(\mathbf{f}'\left(\mathbf{x}\right)\boldsymbol{\theta} + \psi\left(\mathbf{x}\right)\right).$$

Define $\hat{\boldsymbol{\psi}} = \left(\psi\left(\mathbf{x}_1\right), ..., \psi\left(\mathbf{x}_n\right)\right)'$ (here we use all n, not necessarily distinct, points at which observations are made). Conditional on $\boldsymbol{\theta}, \sigma^2$, and using the fact that $\beta_0(\mathbf{x}) = \left(\mathbf{f}'\left(\mathbf{x}\right) - \boldsymbol{\alpha}_0'(\mathbf{x})\,\mathbf{X}\right)\boldsymbol{\theta}_0$, we obtain

$$\begin{aligned} E\left[Z^2(\mathbf{x})|\boldsymbol{\theta}, \sigma^2\right] &= \mathrm{VAR}\left[Z(\mathbf{x})|\boldsymbol{\theta}, \sigma^2\right] + \left\{E\left[Z(\mathbf{x})|\boldsymbol{\theta}, \sigma^2\right]\right\}^2 \\ &= \sigma^2\boldsymbol{\alpha}_0'(\mathbf{x})\boldsymbol{\alpha}_0\left(\mathbf{x}\right) + \left\{\boldsymbol{\alpha}_0'(\mathbf{x})\left(\mathbf{X}\boldsymbol{\theta} + \hat{\boldsymbol{\psi}}\right) + \beta_0(\mathbf{x}) - \left(\mathbf{f}'\left(\mathbf{x}\right)\boldsymbol{\theta} + \psi\left(\mathbf{x}\right)\right)\right\}^2 \\ &= \sigma^2\boldsymbol{\alpha}_0'(\mathbf{x})\boldsymbol{\alpha}_0\left(\mathbf{x}\right) + \left\{\left(\boldsymbol{\alpha}_0'(\mathbf{x})\,\mathbf{X} - \mathbf{f}'\left(\mathbf{x}\right)\right)\left(\boldsymbol{\theta} - \boldsymbol{\theta}_0\right) + \left(\boldsymbol{\alpha}_0'(\mathbf{x})\hat{\boldsymbol{\psi}} - \psi\left(\mathbf{x}\right)\right)\right\}^2. \end{aligned}$$

Upon averaging over $\boldsymbol{\theta}$, $E\left[Z^2(\mathbf{x})|\sigma^2\right]$ becomes $\sigma^2\boldsymbol{\alpha}_0'(\mathbf{x})\boldsymbol{\alpha}_0\left(\mathbf{x}\right) +$

$$\begin{aligned} & E\left[\left\{\left(\boldsymbol{\alpha}_0'(\mathbf{x})\,\mathbf{X} - \mathbf{f}'\left(\mathbf{x}\right)\right)\left(\boldsymbol{\theta} - \boldsymbol{\theta}_0\right) + \left(\boldsymbol{\alpha}_0'(\mathbf{x})\hat{\boldsymbol{\psi}} - \psi\left(\mathbf{x}\right)\right)\right\}^2 |\sigma^2\right] \\ = & \mathrm{VAR}\left[\left(\boldsymbol{\alpha}_0'(\mathbf{x})\,\mathbf{X} - \mathbf{f}'\left(\mathbf{x}\right)\right)\left(\boldsymbol{\theta} - \boldsymbol{\theta}_0\right) + \left(\boldsymbol{\alpha}_0'(\mathbf{x})\hat{\boldsymbol{\psi}} - \psi\left(\mathbf{x}\right)\right)|\sigma^2\right] \\ & + \left\{E\left[\left(\boldsymbol{\alpha}_0'(\mathbf{x})\,\mathbf{X} - \mathbf{f}'\left(\mathbf{x}\right)\right)\left(\boldsymbol{\theta} - \boldsymbol{\theta}_0\right) + \left(\boldsymbol{\alpha}_0'(\mathbf{x})\hat{\boldsymbol{\psi}} - \psi\left(\mathbf{x}\right)\right)|\sigma^2\right]\right\}^2 \\ = & \frac{\sigma^2}{n_0}\left(\boldsymbol{\alpha}_0'(\mathbf{x})\,\mathbf{X} - \mathbf{f}'\left(\mathbf{x}\right)\right)\mathbf{C}_0\left(\mathbf{X}'\boldsymbol{\alpha}_0\left(\mathbf{x}\right) - \mathbf{f}\left(\mathbf{x}\right)\right) + \left(\boldsymbol{\alpha}_0'(\mathbf{x})\hat{\boldsymbol{\psi}} - \psi\left(\mathbf{x}\right)\right)^2, \end{aligned}$$

and then averaging over σ^2 gives

$$\begin{aligned}
\text{MSE}\left(\mathbf{x};\psi\right) &= \sigma_0^2\boldsymbol{\alpha}_0'(\mathbf{x})\boldsymbol{\alpha}_0\left(\mathbf{x}\right) \\
&\quad+\frac{\sigma_0^2}{n_0}\left(\boldsymbol{\alpha}_0'(\mathbf{x})\,\mathbf{X}-\mathbf{f}'\left(\mathbf{x}\right)\right)\mathbf{C}_0\left(\mathbf{X}'\boldsymbol{\alpha}_0\left(\mathbf{x}\right)-\mathbf{f}\left(\mathbf{x}\right)\right)+\left(\boldsymbol{\alpha}_0'(\mathbf{x})\hat{\boldsymbol{\psi}}-\psi\left(\mathbf{x}\right)\right)^2 \\
&= \sigma_0^2\mathbf{f}'\left(\mathbf{x}\right)\left(n_0\mathbf{C}_0^{-1}+\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{f}\left(\mathbf{x}\right)+\left(\boldsymbol{\alpha}_0'(\mathbf{x})\hat{\boldsymbol{\psi}}-\psi\left(\mathbf{x}\right)\right)^2.
\end{aligned}$$

As a check we note that

$$\text{MSE}\left(\mathbf{x};\psi\right)=\text{MSE}_0\left(\boldsymbol{\alpha}_0(\mathbf{x}),\beta_0(\mathbf{x})\right)+\left(\boldsymbol{\alpha}_0'(\mathbf{x})\hat{\boldsymbol{\psi}}-\psi\left(\mathbf{x}\right)\right)^2.$$

FInally, we note that $\mathbf{X}'\mathbf{X}=n\mathbf{M}$ and that

$$\boldsymbol{\alpha}_0'(\mathbf{x})\hat{\boldsymbol{\psi}}=\hat{\boldsymbol{\psi}}'\boldsymbol{\alpha}_0(\mathbf{x})=\hat{\boldsymbol{\psi}}'\mathbf{X}\left(n_0\mathbf{C}_0^{-1}+n\mathbf{M}\right)^{-1}\mathbf{f}\left(\mathbf{x}\right)=n\mathbf{b}'\left(n_0\mathbf{C}_0^{-1}+n\mathbf{M}\right)^{-1}\mathbf{f}\left(\mathbf{x}\right).$$

□

Acknowledgements

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

References

Pukelsheim, F. (1993), *Optimal Design of Experiments*, Wiley.

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