Notes on: Design and M-estimation
Douglas P. Wiens[1]

July 2, 2019

Abstract: Suppose one will estimate the regression parameters using ordinnary M-estimation, with 'M-loss', defined at (5), replacing squared error loss. Although some details remain to be filled in here, it seems to show that, ignoring asymptotically neglible terms, the designs, robust against model misspecification as well as outlying y-values, are *identical* to the minimax robust designs for Least Squares.

1 Introduction

1.1 Model development

We consider an approximate regression model

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

here the values of $\boldsymbol{x} \in \boldsymbol{\chi}$ (the design space) will be chosen by the experimenter. At such values of $\boldsymbol{x}$, $E\left[Y\left(\boldsymbol{x}\right)\right]$ is observed with additive random error:

$$Y\left(\boldsymbol{x}\right) = E\left[Y\left(\boldsymbol{x}\right)\right] + \varepsilon\left(\boldsymbol{x}\right),$$

for i.i.d. errors ε (implying that the distribution of $\varepsilon\left(\boldsymbol{x}\right)$ does not depend on $\boldsymbol{x}$).

The measure $\mu\left(d\boldsymbol{x}\right)$ on $\boldsymbol{\chi} \subset \mathbb{R}^q$ is either Lebesgue measure or counting measure (normed to have unit mass), depending on the structure of the design space. For a convex, even, non-negative function ρ, with derivative ψ, we define $\boldsymbol{\theta}$ to be that which makes makes the approximation in (1) most accurate:

$$\boldsymbol{\theta} = \arg\min_{\mathbf{t}} \int_{\boldsymbol{\chi}} \rho\left(E\left[Y\left(\boldsymbol{x}\right)\right] - \boldsymbol{f}'\left(\boldsymbol{x}\right)\mathbf{t}\right)\mu\left(d\boldsymbol{x}\right).$$

Equivalently, with

$$\tau\left(\boldsymbol{x}\right) \overset{def}{=} E\left[Y\left(\boldsymbol{x}\right)\right] - \boldsymbol{f}'\left(\boldsymbol{x}\right)\boldsymbol{\theta},$$

we have

$$\int_{\boldsymbol{\chi}} \psi\left(\tau\left(\boldsymbol{x}\right)\right)\boldsymbol{f}\left(\boldsymbol{x}\right)\mu\left(d\boldsymbol{x}\right) = \mathbf{0}. \tag{2}$$

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

We bound the approximation error in (1) by assuming that

$$\int_{\boldsymbol{\chi}} \tau^2(\boldsymbol{x})\,\mu(d\boldsymbol{x}) \le \tau_0^2/n, \tag{3}$$

for a user-specified constant τ_0. Here n is the proposed number of observations.

Let ξ_n be the design measure, placing mass

$$\xi_n(\boldsymbol{x}_i) = \frac{n_i}{n}$$

at $\boldsymbol{x}_i$; we may later broaden this definition so as to allow any probability measure on $\boldsymbol{\chi}$ to be a design, with an approximation (so that $n\xi_n(\boldsymbol{x})$ is an integer) possibly required prior to implementation.

Estimation will be by M-estimation:

$$\hat{\boldsymbol{\theta}}_n = \arg\min_{\boldsymbol{\theta}} \int_{\boldsymbol{\chi}} \rho\left(Y(\boldsymbol{x}) - \boldsymbol{f}'(\boldsymbol{x})\,\boldsymbol{\theta}\right)\xi_n(d\boldsymbol{x});$$

equivalently, with residuals

$$\hat{\tau}_n(\boldsymbol{x}) = Y(\boldsymbol{x}) - \boldsymbol{f}'(\boldsymbol{x})\,\hat{\boldsymbol{\theta}}_n$$

we have

$$\int_{\boldsymbol{\chi}} \psi\left(\hat{\tau}_n(\boldsymbol{x})\right)\boldsymbol{f}(\boldsymbol{x})\,\xi_n(d\boldsymbol{x}) = \mathbf{0}. \tag{4}$$

Let Υ be the class of functions $\tau(\cdot)$ defined by (2) and (3). With $\hat{Y}(\boldsymbol{x}) = \boldsymbol{f}'(\boldsymbol{x})\,\hat{\boldsymbol{\theta}}_n$ we define our loss as 'Integrated Expected Error':

$$\begin{aligned} \mathrm{IE}(\xi,\tau) &= \int_{\boldsymbol{\chi}} E\left\{\rho\left(E\left[Y(\boldsymbol{x})\right] - \hat{Y}(\boldsymbol{x})\right)\right\}\mu(d\boldsymbol{x}) \\ &= \int_{\boldsymbol{\chi}} E\left\{\rho\left(\tau(\boldsymbol{x}) - \boldsymbol{f}'(\boldsymbol{x})\left(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}\right)\right)\right\}\mu(d\boldsymbol{x}). \end{aligned} \tag{5}$$

We seek a minimax design ξ_0, i.e. one which minimizes

$$\mathcal{L}(\xi) = \max_{\tau\in\Upsilon} \mathrm{IE}(\xi,\tau).$$

1.2 Asymptotic normality

To start, I think we should determine a sequence $\{\boldsymbol{\theta}_n\}$ for which $\hat{\boldsymbol{\theta}}_n$ is $\sqrt{n}$-consistent – i.e. $\sqrt{n}\left(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_n\right) = O_P(1)$. See (6) for a plausible conjecture. Note that we can write

$$\hat{\tau}_n(\boldsymbol{x}) = \tau(\boldsymbol{x}) + \varepsilon(\boldsymbol{x}) - \boldsymbol{f}'(\boldsymbol{x})\left(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}\right)$$

and then write (4) as

$$\mathbf{0}=\frac{1}{n}\sum_{\{i|n_i>0\}}\sum_{j=1}^{n_i}\psi\left(\{\tau(\boldsymbol{x}_i)+\varepsilon_j(\boldsymbol{x}_i)-\boldsymbol{f}'(\boldsymbol{x}_i)(\boldsymbol{\theta}_n-\boldsymbol{\theta})\}-\boldsymbol{f}'(\boldsymbol{x}_i)\left(\hat{\boldsymbol{\theta}}_n-\boldsymbol{\theta}_n\right)\right)\boldsymbol{f}(\boldsymbol{x}_i);$$

here $\{\varepsilon_j(\boldsymbol{x}_i)\}$ are the errors associated with the n_i replicates at $\boldsymbol{x}_i$.

For remainders R_i, one term expansions around $\boldsymbol{f}'(\boldsymbol{x}_i)\left(\hat{\boldsymbol{\theta}}_n-\boldsymbol{\theta}_n\right)$, i.e.

$$\begin{aligned}
&\psi\left(\{\tau(\boldsymbol{x}_i)+\varepsilon_j(\boldsymbol{x}_i)-\boldsymbol{f}'(\boldsymbol{x}_i)(\boldsymbol{\theta}_n-\boldsymbol{\theta})\}-\boldsymbol{f}'(\boldsymbol{x}_i)\left(\hat{\boldsymbol{\theta}}_n-\boldsymbol{\theta}_n\right)\right)\\
=\;&\psi\left(\{\tau(\boldsymbol{x}_i)+\varepsilon_j(\boldsymbol{x}_i)-\boldsymbol{f}'(\boldsymbol{x}_i)(\boldsymbol{\theta}_n-\boldsymbol{\theta})\}\right)\\
&-\psi'\left(\{\tau(\boldsymbol{x}_i)+\varepsilon_j(\boldsymbol{x}_i)-\boldsymbol{f}'(\boldsymbol{x}_i)(\boldsymbol{\theta}_n-\boldsymbol{\theta})\}\right)\boldsymbol{f}'(\boldsymbol{x}_i)\left(\hat{\boldsymbol{\theta}}_n-\boldsymbol{\theta}_n\right)+R_i
\end{aligned}$$

together with $R_n=\frac{1}{n}\sum_i n_iR_i$ give

$$\begin{aligned}
\mathbf{0}\;=\;&\frac{1}{n}\sum_{\{i|n_i>0\}}\sum_{j=1}^{n_i}\psi\left(\{\tau(\boldsymbol{x}_i)+\varepsilon_j(\boldsymbol{x}_i)-\boldsymbol{f}'(\boldsymbol{x}_i)(\boldsymbol{\theta}_n-\boldsymbol{\theta})\}\right)\boldsymbol{f}(\boldsymbol{x}_i)\\
&-\frac{1}{n}\sum_{\{i|n_i>0\}}\sum_{j=1}^{n_i}\psi'\left(\{\tau(\boldsymbol{x}_i)+\varepsilon_j(\boldsymbol{x}_i)-\boldsymbol{f}'(\boldsymbol{x}_i)(\boldsymbol{\theta}_n-\boldsymbol{\theta})\}\right)\boldsymbol{f}(\boldsymbol{x}_i)\boldsymbol{f}'(\boldsymbol{x}_i)\left(\hat{\boldsymbol{\theta}}_n-\boldsymbol{\theta}_n\right)+R_n,
\end{aligned}$$

whence

$$\begin{aligned}
\sqrt{n}\left(\hat{\boldsymbol{\theta}}_n-\boldsymbol{\theta}_n\right)\;=\;&\left[\frac{1}{n}\sum_{\{i|n_i>0\}}\sum_{j=1}^{n_i}\psi'\left(\{\tau(\boldsymbol{x}_i)+\varepsilon_j(\boldsymbol{x}_i)-\boldsymbol{f}'(\boldsymbol{x}_i)(\boldsymbol{\theta}_n-\boldsymbol{\theta})\}\right)\boldsymbol{f}(\boldsymbol{x}_i)\boldsymbol{f}'(\boldsymbol{x}_i)\right]^{-1}\cdot\\
&\left\{\frac{1}{\sqrt{n}}\sum_{\{i|n_i>0\}}\sum_{j=1}^{n_i}\psi\left(\{\tau(\boldsymbol{x}_i)+\varepsilon_j(\boldsymbol{x}_i)-\boldsymbol{f}'(\boldsymbol{x}_i)(\boldsymbol{\theta}_n-\boldsymbol{\theta})\}\right)\boldsymbol{f}(\boldsymbol{x}_i)+\sqrt{n}R_n\right\}\\
=\;&\left[\int_{\boldsymbol{\chi}}\psi'\left(\{\tau(\boldsymbol{x})+\varepsilon(\boldsymbol{x})-\boldsymbol{f}'(\boldsymbol{x})(\boldsymbol{\theta}_n-\boldsymbol{\theta})\}\right)\boldsymbol{f}(\boldsymbol{x})\boldsymbol{f}'(\boldsymbol{x})\,\xi_n(d\boldsymbol{x})\right]^{-1}\cdot\\
&\cdot\left\{\sqrt{n}\int_{\boldsymbol{\chi}}\psi\left(\{\tau(\boldsymbol{x})+\varepsilon(\boldsymbol{x})-\boldsymbol{f}'(\boldsymbol{x})(\boldsymbol{\theta}_n-\boldsymbol{\theta})\}\right)\boldsymbol{f}(\boldsymbol{x})\,\xi_n(d\boldsymbol{x})+\sqrt{n}R_n\right\}.
\end{aligned}$$

IF we can show that $\sqrt{n}R_n=o_P(1)$ and that $(\boldsymbol{\theta}_n-\boldsymbol{\theta})$ is $O\left(n^{-1/2}\right)$ then this will give an asymptotic normal approximation to the distribution of $\sqrt{n}\left(\hat{\boldsymbol{\theta}}_n-\boldsymbol{\theta}_n\right)$.

Can $\boldsymbol{\theta}_n-\boldsymbol{\theta}$ be defined through (4) and Fisher consistency? This would give

$$\begin{aligned}
\mathbf{0} &= \int_{\boldsymbol{\chi}} E\left[\psi\left(Y\left(\boldsymbol{x}\right)-\boldsymbol{f}'\left(\boldsymbol{x}\right)\boldsymbol{\theta}_n\right)\right]\boldsymbol{f}\left(\boldsymbol{x}\right)\xi_n\left(d\boldsymbol{x}\right)\\
&= \int_{\boldsymbol{\chi}} E\left[\psi\left(\tau\left(\boldsymbol{x}\right)+\varepsilon\left(\boldsymbol{x}\right)-\boldsymbol{f}'\left(\boldsymbol{x}\right)\left(\boldsymbol{\theta}_n-\boldsymbol{\theta}\right)\right)\right]\boldsymbol{f}\left(\boldsymbol{x}\right)\xi_n\left(d\boldsymbol{x}\right)\\
&= \sum_{\{i|n_i>0\}}\sum_{j=1}^{n_i} E\left[\psi\left(\tau\left(\boldsymbol{x}_i\right)+\varepsilon_j\left(\boldsymbol{x}_i\right)-\boldsymbol{f}'\left(\boldsymbol{x}_i\right)\left(\boldsymbol{\theta}_n-\boldsymbol{\theta}\right)\right)\right]\boldsymbol{f}\left(\boldsymbol{x}\right);
\end{aligned}$$

as above this results in

$$\begin{aligned}
\boldsymbol{\theta}_n-\boldsymbol{\theta} &= \left[\frac{1}{n}\sum_{\{i|n_i>0\}}\sum_{j=1}^{n_i} E\psi'\left(\tau\left(\boldsymbol{x}_i\right)+\varepsilon_j\left(\boldsymbol{x}_i\right)\right)\boldsymbol{f}\left(\boldsymbol{x}_i\right)\boldsymbol{f}'\left(\boldsymbol{x}_i\right)\right]^{-1}\cdot\\
&\quad \left\{\frac{1}{n}\sum_{\{i|n_i>0\}}\sum_{j=1}^{n_i} E\left[\psi\left(\tau\left(\boldsymbol{x}_i\right)+\varepsilon_j\left(\boldsymbol{x}_i\right)\right)\right]\boldsymbol{f}\left(\boldsymbol{x}_i\right)+S_n\right\}\\
&= \left[\int_{\boldsymbol{\chi}} E\left[\psi'\left(\tau\left(\boldsymbol{x}\right)+\varepsilon\left(\boldsymbol{x}\right)\right)\right]\boldsymbol{f}\left(\boldsymbol{x}\right)\boldsymbol{f}'\left(\boldsymbol{x}\right)\xi_n\left(d\boldsymbol{x}\right)\right]^{-1}\cdot\\
&\quad \left\{\int_{\boldsymbol{\chi}} E\left[\psi\left(\tau\left(\boldsymbol{x}\right)+\varepsilon\left(\boldsymbol{x}\right)\right)\right]\boldsymbol{f}\left(\boldsymbol{x}\right)\xi_n\left(d\boldsymbol{x}\right)+S_n\right\}. \qquad (6)
\end{aligned}$$

From now on I will assume that the errors are symmetrically distributed, so that $E\left[\psi\left(\varepsilon\right)\right]=E\left[\psi''\left(\varepsilon\right)\right]=0$. It remains to show that the remainder S_n is $o\left(n^{-1/2}\right)$.

If all this can be justified then, ignoring terms which are $o\left(n^{-1/2}\right)$ we have $\boldsymbol{\theta}_n-\boldsymbol{\theta}=\mathbf{M}_n^{-1}\mathbf{b}_n$, for

$$\begin{aligned}
\mathbf{M}_n &= \int_{\boldsymbol{\chi}} E\left[\psi'\left(\tau\left(\boldsymbol{x}\right)+\varepsilon\left(\boldsymbol{x}\right)\right)\right]\boldsymbol{f}\left(\boldsymbol{x}\right)\boldsymbol{f}'\left(\boldsymbol{x}\right)\xi_n\left(d\boldsymbol{x}\right),\\
\mathbf{b}_n &= \int_{\boldsymbol{\chi}} E\left[\psi\left(\tau\left(\boldsymbol{x}\right)+\varepsilon\left(\boldsymbol{x}\right)\right)\right]\boldsymbol{f}\left(\boldsymbol{x}\right)\xi_n\left(d\boldsymbol{x}\right);
\end{aligned}$$

then

$$\begin{aligned}
\sqrt{n}\left(\hat{\boldsymbol{\theta}}_n-\boldsymbol{\theta}_n\right) &= \left[\int_{\boldsymbol{\chi}} \psi'\left(\left\{\tau\left(\boldsymbol{x}\right)+\varepsilon\left(\boldsymbol{x}\right)-\boldsymbol{f}'\left(\boldsymbol{x}\right)\mathbf{M}_n^{-1}\mathbf{b}_n\right\}\right)\boldsymbol{f}\left(\boldsymbol{x}\right)\boldsymbol{f}'\left(\boldsymbol{x}\right)\xi_n\left(d\boldsymbol{x}\right)\right]^{-1}\cdot\\
&\quad \left\{\sqrt{n}\int_{\boldsymbol{\chi}} \psi\left(\left\{\tau\left(\boldsymbol{x}\right)+\varepsilon\left(\boldsymbol{x}\right)-\boldsymbol{f}'\left(\boldsymbol{x}\right)\mathbf{M}_n^{-1}\mathbf{b}_n\right\}\right)\boldsymbol{f}\left(\boldsymbol{x}\right)\xi_n\left(d\boldsymbol{x}\right)+o_P(1)\right\}\\
&\stackrel{def}{=} \mathbf{A}_n^{-1}\left\{\sqrt{n}\mathbf{z}_n+o_P(1)\right\}.
\end{aligned}$$

Thus

$$\sqrt{n}\left(\hat{\boldsymbol{\theta}}_n-\boldsymbol{\theta}\right)\sim\sqrt{n}\left(\boldsymbol{\theta}_n-\boldsymbol{\theta}\right)+\sqrt{n}\mathbf{A}_n^{-1}\mathbf{z}_n=\sqrt{n}\mathbf{M}_n^{-1}\mathbf{b}_n+\sqrt{n}\mathbf{A}_n^{-1}\mathbf{z}_n. \qquad (7)$$

1.3 Evaluation of IE

To evaluate (6), and then (5), this, first expand $\mathbf{M}_n$ and $\mathbf{b}_n$ around $\tau(\boldsymbol{x})$ as

$$\begin{aligned}
\mathbf{M}_n &= \int_{\boldsymbol{\chi}} E\left[\psi'(\varepsilon(\boldsymbol{x})) + \psi''(\varepsilon(\boldsymbol{x}))\tau(\boldsymbol{x})\right]\boldsymbol{f}(\boldsymbol{x})\boldsymbol{f}'(\boldsymbol{x})\xi_n(d\boldsymbol{x}) + o\left(n^{-1/2}\right) \\
&= E\left[\psi'(\varepsilon)\right]\int_{\boldsymbol{\chi}} \boldsymbol{f}(\boldsymbol{x})\boldsymbol{f}'(\boldsymbol{x})\xi_n(d\boldsymbol{x}) + o\left(n^{-1/2}\right), \\
\mathbf{b}_n &= \int_{\boldsymbol{\chi}} E\left[\psi(\varepsilon(\boldsymbol{x})) + \psi'(\varepsilon(\boldsymbol{x}))\tau(\boldsymbol{x})\right]\boldsymbol{f}(\boldsymbol{x})\xi_n(d\boldsymbol{x}) + o\left(n^{-1/2}\right) \\
&= E\left[\psi'(\varepsilon)\right]\int_{\boldsymbol{\chi}} \boldsymbol{f}(\boldsymbol{x})\tau(\boldsymbol{x})\xi_n(d\boldsymbol{x}) + o\left(n^{-1/2}\right);
\end{aligned}$$

thus

$$\mathbf{M}_n^{-1}\mathbf{b}_n = \mathbf{M}_{0,n}^{-1}\mathbf{b}_{0,n} + o\left(n^{-1/2}\right), \tag{8}$$

where

$$\mathbf{M}_{0,n} = \int_{\boldsymbol{\chi}} \boldsymbol{f}(\boldsymbol{x})\boldsymbol{f}'(\boldsymbol{x})\xi_n(d\boldsymbol{x}), \qquad \mathbf{b}_{0,n} = \int_{\boldsymbol{\chi}} \boldsymbol{f}(\boldsymbol{x})\tau(\boldsymbol{x})\xi_n(d\boldsymbol{x}).$$

Now note that by the WLLN, $\mathbf{A}_n = E[\mathbf{A}_n] + o_p(1)$, where, expanding around $\tau(\boldsymbol{x}) - \boldsymbol{f}'(\boldsymbol{x})\mathbf{M}_{0,n}^{-1}\mathbf{b}_{0,n}$,

$$\begin{aligned}
E[\mathbf{A}_n] &= \int_{\boldsymbol{\chi}} E\left[\psi'\left(\left\{\tau(\boldsymbol{x}) + \varepsilon(\boldsymbol{x}) - \boldsymbol{f}'(\boldsymbol{x})\mathbf{M}_{0,n}^{-1}\mathbf{b}_{0,n}\right\}\right)\right]\boldsymbol{f}(\boldsymbol{x})\boldsymbol{f}'(\boldsymbol{x})\xi_n(d\boldsymbol{x}) + o\left(n^{-1/2}\right) \\
&= \int_{\boldsymbol{\chi}} \left[\psi'(\varepsilon(\boldsymbol{x})) + \psi''(\varepsilon(\boldsymbol{x}))\left(\tau(\boldsymbol{x}) - \boldsymbol{f}'(\boldsymbol{x})\mathbf{M}_{0,n}^{-1}\mathbf{b}_{0,n}\right)\right]\boldsymbol{f}(\boldsymbol{x})\boldsymbol{f}'(\boldsymbol{x})\xi_n(d\boldsymbol{x}) + o\left(n^{-1/2}\right) \\
&= \mathbf{M}_n + O\left(n^{-1/2}\right).
\end{aligned}$$

Finally,

$$\begin{aligned}
\sqrt{n}\mathbf{z}_n &= \sqrt{n}\int_{\boldsymbol{\chi}} \psi\left(\left\{\tau(\boldsymbol{x}) + \varepsilon(\boldsymbol{x}) - \boldsymbol{f}'(\boldsymbol{x})\mathbf{M}_n^{-1}\mathbf{b}_n\right\}\right)\boldsymbol{f}(\boldsymbol{x})\xi_n(d\boldsymbol{x}) \\
&= \sqrt{n}\int_{\boldsymbol{\chi}} \left[\psi(\varepsilon(\boldsymbol{x})) + \psi'(\varepsilon(\boldsymbol{x}))\left(\tau(\boldsymbol{x}) - \boldsymbol{f}'(\boldsymbol{x})\mathbf{M}_n^{-1}\mathbf{b}_n\right)\right]\boldsymbol{f}(\boldsymbol{x})\xi_n(d\boldsymbol{x}) + o_p(1)
\end{aligned}$$

where it should be easy to show that

$$\begin{aligned}
&\sqrt{n}\int_{\boldsymbol{\chi}} \psi'(\varepsilon(\boldsymbol{x}))\left(\tau(\boldsymbol{x}) - \boldsymbol{f}'(\boldsymbol{x})\mathbf{M}_n^{-1}\mathbf{b}_n\right)\boldsymbol{f}(\boldsymbol{x})\xi_n(d\boldsymbol{x}) \\
&= \int_{\boldsymbol{\chi}} E\left[\psi'(\varepsilon(\boldsymbol{x}))\sqrt{n}\left(\tau(\boldsymbol{x}) - \boldsymbol{f}'(\boldsymbol{x})\mathbf{M}_n^{-1}\mathbf{b}_n\right)\right]\boldsymbol{f}(\boldsymbol{x})\xi_n(d\boldsymbol{x}) + o_p(1) \\
&= E\left[\psi'(\varepsilon)\right]\int_{\boldsymbol{\chi}} \sqrt{n}\tau(\boldsymbol{x})\boldsymbol{f}(\boldsymbol{x})\xi_n(d\boldsymbol{x}) - E\left[\psi'(\varepsilon)\right]\sqrt{n}\int_{\boldsymbol{\chi}} \boldsymbol{f}'(\boldsymbol{x})\boldsymbol{f}(\boldsymbol{x})\xi_n(d\boldsymbol{x})\mathbf{M}_n^{-1}\mathbf{b}_n + o_p(1) \\
&= \sqrt{n}\mathbf{b}_n - \sqrt{n}\mathbf{M}_n\mathbf{M}_n^{-1}\mathbf{b}_n + o_p(1),
\end{aligned}$$

which is $o_p(1)$. Thus

$$\begin{aligned}\sqrt{n}\mathbf{z}_n &= \sqrt{n}\int_{\boldsymbol{\chi}} \psi\left(\varepsilon\left(\boldsymbol{x}\right)\right)\boldsymbol{f}\left(\boldsymbol{x}\right)\xi_n\left(d\boldsymbol{x}\right) + o_p\left(1\right) \\ &= \frac{1}{\sqrt{n}}\sum_{\{i|n_i>0\}}\sum_{j=1}^{n_i}\psi\left(\varepsilon_j\left(\boldsymbol{x}_i\right)\right)\boldsymbol{f}\left(\boldsymbol{x}_i\right) + o_p\left(1\right) \\ &\sim AN\left(\mathbf{0}, E\left[\psi^2\left(\varepsilon\right)\right]\mathbf{M}_{0,n}\right),\end{aligned}$$

and so

$$\sqrt{n}\mathbf{A}_n^{-1}\mathbf{z}_n \sim AN\left(\mathbf{0}, \frac{E\left[\psi^2\left(\varepsilon\right)\right]}{\left(E\left[\psi'\left(\varepsilon\right)\right]\right)^2}\mathbf{M}_{0,n}^{-1}\right). \tag{9}$$

By (7), (8) and (9), and with $\sigma_\varepsilon^2 \stackrel{def}{=} E\left[\psi^2\left(\varepsilon\right)\right]/\left(E\left[\psi'\left(\varepsilon\right)\right]\right)^2$,

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

Expanding (5) and using (2) results in

$$\begin{aligned}IE &= \int_{\boldsymbol{\chi}}\rho\left(\tau\left(\boldsymbol{x}\right)\right)\mu\left(d\boldsymbol{x}\right) + \psi'\left(0\right)\int_{\boldsymbol{\chi}}\boldsymbol{f}'\left(\boldsymbol{x}\right)E\left[\left(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}\right)\left(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}\right)'\right]\boldsymbol{f}\left(\boldsymbol{x}\right)\mu\left(d\boldsymbol{x}\right) + o\left(n^{-1}\right) \\ &= \int_{\boldsymbol{\chi}}\rho\left(\tau\left(\boldsymbol{x}\right)\right)\mu\left(d\boldsymbol{x}\right) + \psi'\left(0\right)\int_{\boldsymbol{\chi}}\boldsymbol{f}'\left(\boldsymbol{x}\right)\left[\mathbf{M}_{0,n}^{-1}\mathbf{b}_{0,n}\mathbf{b}_{0,n}'\mathbf{M}_{0,n}^{-1} + \frac{\sigma_\varepsilon^2}{n}\mathbf{M}_{0,n}^{-1}\right]\boldsymbol{f}\left(\boldsymbol{x}\right)\mu\left(d\boldsymbol{x}\right) + o\left(n^{-1}\right) \\ &= \int_{\boldsymbol{\chi}}\rho\left(\tau\left(\boldsymbol{x}\right)\right)\mu\left(d\boldsymbol{x}\right) + \psi'\left(0\right)\mathbf{b}_{0,n}'\mathbf{M}_{0,n}^{-1}\mathbf{C}\mathbf{M}_{0,n}^{-1}\mathbf{b}_{0,n}' + \frac{\sigma_\varepsilon^2}{n}tr\mathbf{C}\mathbf{M}_{0,n}^{-1} + o\left(n^{-1}\right),\end{aligned} \tag{10}$$

where

$$\mathbf{C} = \int_{\boldsymbol{\chi}}\boldsymbol{f}\left(\boldsymbol{x}\right)\boldsymbol{f}'\left(\boldsymbol{x}\right)\mu\left(d\boldsymbol{x}\right).$$

2 Summary and conclusions

Ignoring terms which are $o\left(n^{-1}\right)$, the loss is given by (10). This is exact when $\rho\left(\tau\right) = \tau^2$, i.e. for LSE, and in general differs from that for LSE only in the additive term $\int_{\boldsymbol{\chi}}\rho\left(\tau\left(\boldsymbol{x}\right)\right)\mu\left(d\boldsymbol{x}\right)$, which does not depend on the design, and the multiplicative constant $\psi'\left(0\right)$, which can be absorbed into σ_ε^2. Thus the ensuing minimax design problem is EXACTLY the same as for least squares – i.e., is a problem which has already been solved. Too bad.