This work preceded, and evolved into, Wiens (2003). It utilizes a different neighbourhood structure, to formalise departures from the nominal regression model, than was used there.

AMS 1991 SUBJECT CLASSIFICATIONS: Primary 62K05, 62F35, 92C50; secondary 62J05.

KEY WORDS: M-estimation; Minimax; Optimal design.
1. INTRODUCTION

In this article we pose a generalisation of the following problem of interest in clinical trials, and study its solution. Suppose that \( n \) subjects enter a clinical study and are to be assigned to one of two treatment groups, referred to here as ‘treatment’ and ‘control’. Corresponding to each subject is a vector \( \mathbf{x} \) of prognostic factors or covariates, upon which the mean response relies through a vector \( \mathbf{z}(\mathbf{x}) \) of possible regressors. Upon observing \( \mathbf{x} \), the experimenter is to assign the subject to one of the groups. The aim is to obtain an efficient and robust estimate of the difference in the mean responses to the treatments.

A complicating factor calling for a robust solution is that the fitted model relating the response to treatment/covariate effects may be only approximately valid. We entertain two such approximate models:

\[
\text{Model 1}^0: \quad Y = \theta_1 u_1 + \theta_2 u_2 + n^{-1/2} f_u(\mathbf{x}) + \sigma_u \varepsilon,
\]
\[
\text{Model 2}^0: \quad Y = \theta_1 u_1 + \theta_2 u_2 + \mathbf{z}^T(\mathbf{x}) \phi + n^{-1/2} f_u(\mathbf{x}) + \sigma_u \varepsilon.
\]

In each case \( \mathbf{u} = (u_1, u_2)^T = (0, 1)^T \) for the treatment group, \( = (1, 0)^T \) for the control group and \( \varepsilon \) denotes random error with zero mean and unit variance. The function \( f_u(\mathbf{x}) \) is unknown to the experimenter and serves to formalise the approximate nature of the fitted models, which are \( \hat{Y} = \hat{\theta}_1 u_1 + \hat{\theta}_2 u_2 \) and \( \hat{Y} = \hat{\theta}_1 u_1 + \hat{\theta}_2 u_2 + \mathbf{z}^T(\mathbf{x}) \hat{\phi} \) respectively. In either case we are interested primarily in the parameter \( \theta_2 - \theta_1 \).

The presence of \( f_u(\mathbf{x}) \) implies that the least squares estimate \( \hat{\theta}_2 - \hat{\theta}_1 \) is biased, so that an optimal assignment of subjects to groups should aim at the minimisation of the mean squared error (\( mse \)) rather than merely the variance. Since \( f_u(\cdot) \) is unknown, an appealing approach is to assign it a neighbourhood structure and obtain minimax assignment procedures. These minimise the maximum value of the \( mse \), with the maximum evaluated as \( f \) varies over its neighbourhood.

The purpose of the \( n^{-1/2} \) in the models, which has no effect in finite samples since it can be absorbed by \( f \), is to ensure that bias and variance are of the same order asymptotically.

Ethical considerations often dictate that there should be some randomness in the assignment of subjects to treatments. Thus, we shall derive a probability \( \rho(\mathbf{x}) \) according to which a subject exhibiting covariates \( \mathbf{x} \) is to be assigned to the treatment group. The actual assignment is then made after carrying out a Bernoulli trial with success probability \( \rho(\mathbf{x}) \).
In Section 2 we present a generalisation of this problem to \( p \geq 2 \) treatment groups, and give a formal definition of the neighbourhood structure referred to above. We present an asymptotic approximation to the resulting mean squared error (mse), maximised over the surrounding neighbourhood. This maximum is rather unwieldy except in the special case of the problem outlined above with \( p = 2 \). Thus, we also give a more tractable upper bound on this quantity.

In Section 3 we give an exact solution to the “\( p = 2 \)” problem. It turns out that the solution for Model 2 is, under certain assumptions on the structure of the regressors and the space of covariates, identical to the solution for Model 1. In Section 4 we study assignment probabilities \( \rho(x) \) which minimise the upper bound on the maximum \( mse \), for \( p \geq 2 \).

Atkinson (1982) obtained D-optimal designs for problems similar to those considered here, but assuming the fitted models to be exactly correct. Thus variance minimisation was the only concern. Heckman (1987) considered a problem closer in nature to ours. Her approach assumed \( p = 2 \) and a single covariate, i.e. \( z(x) = x \) in the notation above. She assumed a neighbourhood structure somewhat less broad than ours, and obtained allocations which were minimax when employed with estimates \( \hat{\theta} \) which were themselves minimax, for the given neighbourhood structure, within the class of linear estimates. Thus these are not the classical least squares estimates, a feature which may explain why the corresponding allocation methods have not become more popular. In Section 5 we assess our allocations numerically and compare them with those of Atkinson (1982), Heckman (1987) (in conjunction with least squares estimation) and others.

2. A GENERALISED DESIGN PROBLEM

We consider regression models in which a subject with covariates \( x_1, \ldots, x_t \) yields a response \( Y \) satisfying

\[
Y = E[Y|\mathbf{u}, \mathbf{x}] + \sigma_u \varepsilon
\]

where \( \mathbf{u}_{p \times 1} \) is a vector of regressors and \( \mathbf{x} \) is the vector of covariates ranging over a space \( \mathcal{S} \). We assume that \( \mathbf{x} \) has a density \( m(\mathbf{x}) \) on \( \mathcal{S} \). The regression response is a function of \( \mathbf{u} \) and of regressors \( z(\mathbf{x})_{q \times 1} \), although the regressors are not necessarily modelled by the experimenter. We consider two cases. In each,

\[
E[Y|\mathbf{u}, \mathbf{x}] \approx \mathbf{v}^T(\mathbf{x})\psi. \tag{1}
\]
In Model 1, the fitted response is \( \hat{Y} = \mathbf{u}^T \hat{\theta} \) and we take \( \mathbf{v}_u(x) \equiv \mathbf{u} \), \( \psi = \theta_{p \times 1} \). In Model 2 the fitted response is \( \hat{Y} = \mathbf{u}^T \hat{\theta} + \mathbf{z}^T(x) \hat{\phi} \) and we take \( \mathbf{v}_u(x) = (\mathbf{u}^T, \mathbf{z}^T(x))^T \), \( \psi = (\theta^T, \phi^T)^T \) for a vector \( \phi_{q \times 1} \).

In our consideration of Model 2 we shall assume that the covariates occur linearly: \( \mathbf{z}(x) = x \) and that, perhaps after an appropriate transformation, \( S \) is a sphere with centre \( \mathbf{0} \) within which the density \( m(x) \) is symmetric in each coordinate.

**Example 1.1.** Suppose that, after an affine transformation, the covariates \( x \) follow a \( q \)-dimensional unit normal distribution restricted to a sphere \( S \) which is of finite measure, but large enough to contain all of the data with high probability. In practice one could first transform to standardised covariates, *viz.* \( x \to S^{-1/2}(x - \mathbf{t}) \), where \( S \) is a scatter matrix and \( \mathbf{t} \) a location estimate, and then choose a radius at least as large as the maximum value of \( ||x|| \). For this example, to be developed further in subsequent sections, we let \( \phi(x) = (2\pi)^{-q/2} \exp(-||x||^2/2) \) denote the \( N_q(0, \mathbf{I}) \) density and \( H(\cdot) \) the \( \chi^2_q \) distribution function. Then for some \( c \) slightly less than 1 we take

\[
S = \{ \mathbf{x} \ | \ ||\mathbf{x}||^2 \leq h^2 := H^{-1}(c) \},
\]

\[
m(x) = \phi(x)/c, \ x \in S.
\]

Define \( P \) to be \( p \) for Model 1 and \( p + q \) for Model 2. The ‘true’ parameter \( \psi_{P \times 1} \) is defined by

\[
\psi = \arg \min_{\beta} \sum_{\mathbf{u}} \int_S \left\{ E[Y|\mathbf{u}, x] - \mathbf{v}^T_u(x)\beta \right\}^2 dx.
\]

We assume that the approximation (1) is sufficiently accurate that this integral exists for at least some \( \beta \). Now define

\[
f_u(x) = \sqrt{n} \left( E[Y|\mathbf{u}, x] - \mathbf{v}^T_u(x)\psi \right),
\]

so that

\[
Y = \mathbf{v}^T_u(x)\psi + n^{-1/2}f_u(x) + \sigma_u \varepsilon
\]

(2)

with \( \sum_{\mathbf{u}} \int_S f_u(x)v_u(x)dx = 0 \) for each \( \mathbf{u} \). We shall impose the more severe restriction

\[
\int_S f_u(x)v_u(x)dx = 0 \text{ for each } \mathbf{u}.
\]

(3)
In order that errors due to model misspecification not swamp those due to random variation, we shall also assume that

\[ \int_S f_u^2(x) dx \leq \eta_u^2 \quad (4) \]

for given, finite bounds \( \eta_u^2 \).

Suppose now that \( u \) takes on only \( p \) possible values \( u_i = (0, ..., 0, 1, 0, ..., 0)^T \) \( (i = 1, ..., p) \) and that a subject with covariates \( x \) is assigned to a group defined by \( u = u_i \) with probability \( \rho_i(x) = P(i|x) \), independently of all other assignments. Then the joint density of the covariates and probability that a covariate results in an assignment to group \( i \) is

\[ m_i(x) = \rho_i(x)m(x). \]

Put also \( f_i(x) = f_{u_i}(x), v_i(x) = v_{u_i}(x) \) and \( \sigma_i = \sigma_{u_i} \).

The experimenter is assumed to be interested in estimating a complete set of \( p - 1 \) orthonormal contrasts \( W_0 \theta \). We shall derive the mean squared error matrix \( \text{MSE}(W_0 \hat{\theta}) \) of \( W_0 \hat{\theta} \). The loss function is taken to be the almost sure limit of the normalised determinant of this matrix, maximised over the \( f_i \):

\[ \mathcal{L}(\rho_1, ..., \rho_p) = \max_{f_1, ..., f_p} \lim_{n \to \infty} |n \text{MSE}(W_0 \hat{\theta})| \]

with the maxima evaluated subject to (3) and (4).

**Theorem 1.** Assume that

\[ \int_S \|v_i(x)\|^2 m^k(x) dx < \infty \text{ for } k = 0, 1, 2 \text{ and } i = 1, ..., p. \]

Define \( P \times 1 \) vectors and \( P \times P \) matrices

\[ b_i(f_i) = \int_S v_i(x)m_i(x)f_i(x) dx, \]

\[ A_i = \int_S v_i(x)v_i^T(x) dx, \]

\[ B_i = \int_S v_i(x)v_i^T(x)m_i(x) dx, \]

\[ C_i = \int_S v_i(x)v_i^T(x)m_i^2(x) dx, \]

\[ G_i = C_i - B_i A_i^+ B_i; \]
\[ A_i^+ \] denotes the Moore-Penrose generalised inverse) and
\[ B = \sum_{i=1}^{p} B_i, \quad Q = \sum_{i=1}^{p} \sigma_i^2 B_i, \quad b(f_1, ..., f_p) = \sum_{i=1}^{p} b_i(f_i). \quad (5) \]

Define \( W_{p-1 \times p} = W_0 \) for Model 1, \( = \begin{pmatrix} W_0 & 0_{p-1 \times q} \end{pmatrix} \) for Model 2. Let \( R_{p-1 \times p} \), of rank \( p-1 \), be defined in such a way that
\[ B^{-1} W^T (W B^{-1} Q B^{-1} W^T)^{-1} W B^{-1} = R^T R. \quad (6) \]

Then
\[ \mathcal{L}(\rho_1, ..., \rho_p) = |W B^{-1} Q B^{-1} W^T| \left( 1 + \max_{f_1, ..., f_p} ||R b(f_1, ..., f_p)||^2 \right) \quad (7) \]
with
\[ \max_{f_1, ..., f_p} ||R b(f_1, ..., f_p)||^2 = \left( \max_{||\alpha||=1} \sum_{i=1}^{p} \eta_i \sqrt{\alpha^T R G_i R^T \alpha} \right)^2. \quad (8) \]

For \( p > 2 \) the final maximum in (8) is too intractable to admit an exact analysis. We shall replace it with an upper bound which is exact if \( p = 2 \). For this, we first define
\[ a = \int_S dx, \]
\[ s_i = \int_S \rho_i(x) m(x) dx, \quad \quad \quad (9) \]
\[ g_i = \int_S \left( \rho_i(x) m(x) - \frac{s_i}{a} \right)^2 dx. \]

We also define the variance ratios \( \omega_i^2 = \sigma_i^2 / \sigma_1^2 \), the ‘noise/noise’ ratios \( \kappa_i = \eta_i / \eta_1 \) and the ‘noise to contamination’ ratio \( \nu = \sigma_i^2 / \eta_1^2 \). The variance ratios would typically be estimated, perhaps from prior samples or perhaps sequentially as the experiment progresses. The ratios \( \kappa_i \) and \( \nu \) are to be chosen by the experimenter, reflecting his opinion of the relative uncertainties of the two fitted models, and of the relative importance of variance versus bias, respectively. Although our asymptotic approach has eliminated the dependence on the sample size, for large samples we would anticipate a quite small choice of \( \nu \). This is due to the fact that the variance of the estimator is \( O(n^{-1}) \), while the bias is \( O(1) \).

An alternate method of choosing \( \kappa_i \) and/or \( \nu \) is for the experimenter to specify in advance the unconditional probabilities \( s_i \) of assignments to the treatments, and to then choose the parameters so that these \( s_i \) become optimal. See Example 3.1 and Figure 1 for an instance of this approach.
Theorem 2. Define a matrix \( D = \text{diag} (s_1, s_2/\omega_2^2, ..., s_p/\omega_p^2) \). For Model 2, make the restriction to functions \( \rho_i(x) \) which are symmetric in each coordinate. Then for both Model 1 and Model 2

\[
\mathcal{L}(\rho_1, ..., \rho_p) = \frac{\sigma_1^{2(p-1)\text{tr}D}}{p|D|} \left[ 1 + \left\{ \max_{||\alpha||=1} \sum_{i=1}^{p} \eta_i \sqrt{g_i} |\alpha^T R u_i| \right\}^2 \right],
\]

where \( R \) is as defined for Model 1 at (6). The inequality

\[
\mathcal{L}(\rho_1, ..., \rho_p) \leq \frac{\eta^2 \sigma_1^{2(p-2)}}{p} \mathcal{L}'(\rho_1, ..., \rho_p)
\]

holds, where

\[
\mathcal{L}'(\rho_1, ..., \rho_p) = \frac{1}{|D|} \left[ \nu \text{tr}D + \left\{ \sum_{i=1}^{p} \frac{\kappa_i}{\omega_i} \left( \sqrt{g_i} - \frac{\sqrt{\eta_i} \sqrt{g_i}}{s_i} \right) \right\}^2 \right].
\]

In particular, if \( p = 2 \) the upper bound is exact:

\[
2\eta_1^{-2} \mathcal{L}(\rho_1, \rho_2) = \mathcal{L}'(\rho_1, \rho_2) = \nu \left( \frac{1}{s_1} + \frac{\omega_2^2}{s_2} \right) + \left\{ \frac{\sqrt{g_1}}{s_1} + \kappa_2 \frac{\sqrt{g_2}}{s_2} \right\}^2 .
\]

Note that although the loss is formally identical for both models, we do not mean to imply that fitting the covariates is irrelevant. Since the functions \( f_i(x) \) represent the contributions to the response of the covariate information not included in the fitted model, we would expect the \( \eta_i \) to be set at smaller values in Model 2 than in Model 1.

3. EXACT SOLUTIONS WHEN \( p = 2 \)

For the case \( p = 2 \) as described in the Introduction we suppose that \( i = 1 \) denotes the control group, \( i = 2 \) the treatment group. Define \( s = s_2, \omega = \omega_2, \kappa = \kappa_2, \rho = \rho_2 \). Then \( \rho_1 = 1 - \rho \) and we write \( \mathcal{L}(\rho_1, \rho_2) \) as \( \mathcal{L}(\rho) \). Define \( r_1 = \int_S (1 - \rho(x))^2 m^2(x) dx \), \( r_2 = \int_S \rho^2(x)m^2(x) dx \), \( r = \int_S m^2(x) dx \). From Theorem 2,

\[
2\eta_1^{-2} \mathcal{L}(\rho) = \nu \left\{ \frac{1}{1-s} + \frac{\omega^2}{s} \right\} + \left\{ \sqrt{\frac{r_1}{(1-s)^2}} + \frac{1}{\kappa} \sqrt{\frac{r_2}{s^2} - \frac{1}{\alpha}} \right\}^2 ,
\]

to be minimised over functions \( \rho : S \to [0,1] \). The least favourable model contaminants, as brought out in the proof of Theorem 1 - see (A.1) - are

\[
f_i(x) = \frac{\eta_i}{\sqrt{g_i}} (m_i(x) - s_i / a) .
\]

We first minimise (12) for fixed \( s \). For this we require the following result, proven in the Appendix.
Lemma 1. Let \( l_t(x) = (1 - t)l_0(x) + tl_1(x) \), where \( l_0 \) and \( l_1 \) are square integrable functions of \( x \in S \). Then \( \varphi(t) = (\int_S l_t^2(x)dx)^{1/2} \) is a convex function of \( t \in [0, 1] \).

By this Lemma the second term in braces in (12), and hence \( L(p) \), is minimised for fixed \( s \) by a function \( \rho_0 \) iff the convex function
\[
\varphi(t) = \sqrt{\int_S (1 - \rho_t(x))^2 m^2(x)dx} - \frac{1}{a} + \kappa \sqrt{\int_S \rho_t^2(x)m^2(x)dx} - \frac{1}{a} - \varrho \int_S \rho_t(x)m(x)dx
\]
satisfies \( \varphi'(0) \geq 0 \) for all \( \rho_0 : S \rightarrow [0, 1] \), where \( \rho_t(x) = (1 - t)\rho_0(x) + t\rho_\ast(x) \) and \( \varrho \) is a Lagrange multiplier. We calculate that
\[
\varphi'(0) = \int_S (\rho_\ast(x) - \rho_0(x)) m(x) \left\{ (\varrho' + \varrho'') m(x) \rho_0(x) - (\varrho' m(x) + \varrho) \right\} dx
\]
for certain positive constants \( \varrho' \) and \( \varrho'' \) depending on \( \rho_0 \) and \( \varrho \). Explicitly,
\[
\begin{align*}
\varrho' &= \frac{1}{(1 - s)^2} \left( \int_S (1 - \rho_0(x))^2 m^2(x)dx - \frac{1}{a} \right)^{-1/2}, \\
\varrho'' &= \frac{\kappa}{s^2} \left( \int_S \rho_0^2(x)m^2(x)dx - \frac{1}{a} \right)^{-1/2}.
\end{align*}
\]
Thus \( \rho_0 \) is necessarily of the form
\[
\rho_0(x) = \left[ \tau + \frac{\gamma}{m(x)} \right]_0^1
\]
(here\([\cdot]\)_0^1 denotes truncation at 0 and at 1) for \( \gamma = \varrho' / (\varrho' + \varrho'') \) and \( \tau = \varrho' / (\varrho' + \varrho'') \). This last equality is equivalent to
\[
\kappa \tau (1 - s)\sqrt{ar_1(1 - s)^2} = (1 - \tau)s\sqrt{ar_2} - s^2. \tag{13}
\]
The constants \( \gamma \) and \( \tau \) are chosen so that \( \rho_0 \) satisfies (9) and minimises (12).

The minimising choices must and do also satisfy (13), which is a useful check. Indeed, we could equivalently obtain \( \gamma \) and \( \tau \) in terms of \( s \) by solving (9) and (13), and then choosing \( s \) to minimise (12). This approach yields the intuitively pleasing result that if \( \omega = \kappa \), i.e. \( \sigma_2/\sigma_1 = \eta_2/\eta_1 \), then \( \gamma = 0 \), \( \tau = s \) and \( \rho_0(x) \equiv s = \omega / (1 + \omega) \). This choice of \( (\gamma, \tau) \) satisfies (13) and makes the second (bias) term of (12) identically equal to \( (1 + \kappa)^2(r - 1/a) \), independent of \( s \). The first (variance) term is then minimised by \( s = \omega / (1 + \omega) \).
We define these constant assignment probabilities by
\[ \rho_K(x) \equiv \omega/(1 + \omega). \]

The general form of the solution depends on the sign of \( \gamma \):
\[
\rho_0(x) = \begin{cases} 
\min(\tau + \frac{\gamma}{m(x)}, 1) = \begin{cases} 
\tau + \frac{\gamma}{m(x)}, & m(x) \geq \frac{\gamma}{\tau}, \text{ if } \gamma \geq 0; \\
1, & m(x) \leq \frac{\gamma}{\tau},
\end{cases} \\
\max(0, \tau + \frac{\gamma}{m(x)}) = \begin{cases} 
\tau + \frac{\gamma}{m(x)}, & m(x) \geq \frac{|\gamma|}{\tau}, \text{ if } \gamma \leq 0; \\
0, & m(x) \leq \frac{|\gamma|}{\tau},
\end{cases}
\end{cases}
\]

For computing purposes it is simplest to express everything in terms of three functions of \( \gamma \) and \( \tau \):
\[
\xi_k = \xi_k(\gamma, \tau) = \begin{cases} 
\int_{m(x) \geq \gamma/(1-\tau)} m^k(x) dx, & \gamma \geq 0, \\
\int_{m(x) \geq \gamma/\tau} m^k(x) dx, & \gamma \leq 0,
\end{cases} \quad k = 0, 1, 2.
\]

We calculate that then
\[
s = \begin{cases} 
1 - (1 - \tau)\xi_1 + \gamma\xi_0, & \gamma \geq 0, \\
\tau\xi_1 + \gamma\xi_0, & \gamma \leq 0;
\end{cases}
\]
\[
r_1 = \begin{cases} 
(1 - \tau)^2\xi_2 - 2\gamma(1 - \tau)\xi_1 + \gamma^2\xi_0, & \gamma \geq 0, \\
r - (1 - (1 - \tau)^2)\xi_2 - 2\gamma(1 - \tau)\xi_1 + \gamma^2\xi_0, & \gamma \leq 0.
\end{cases}
\]
\[
r_2 = \begin{cases} 
r - (1 - \tau^2)\xi_2 + 2\gamma\tau\xi_1 + \gamma^2\xi_0, & \gamma \geq 0, \\
\tau^2\xi_2 + 2\gamma\tau\xi_1 + \gamma^2\xi_0, & \gamma \leq 0;
\end{cases}
\]

These expressions can now be substituted into (12) and the result minimised numerically over \( \gamma \) and \( \tau \in [0, 1] \), for given choices of \( m(\cdot) \).

For \( q = 1 \) it is possible to compare our results with those of Heckman (1987), who obtained minimax allocation schemes in the case of a single covariate \( x \in [-1, 1] \), as briefly described in §1. In our notation the neighbourhood structure used was as at (2) with \( f(x; u) \) replaced by \( f_i(x_0, x) \) satisfying \( |n^{-1/2}f_i(x_0, x)| \leq \eta_1(x_0)|x - x_0| \) for specified \( x_0 \in [-1, 1] \) and functions \( \eta_1(\cdot) \). The allocation probability \( \rho_H(x) = P(\text{treatment} | x) \) was obtained as
\[
\rho_H(x) = \frac{[\omega^2K(x)]^{1/3}}{1 + [\omega^2K(x)]^{1/3}}
\]
with \( K(x) = \eta_2(x)/\eta_1(x) \) assumed known. Note that \( \rho_H(x) = \rho_K(x) \) if \( \kappa(x) \equiv \omega \).

**Example 3.1.** (Continuation of Example 1.1). Let \( V = \pi^{q/2}/\Gamma\left(\frac{q}{2} + 1\right) \), the volume of the unit sphere in \( \mathbb{R}^q \). Define
\[
d^2 = \begin{cases} 
\min(h^2, \max(0, -2\log((1-\gamma)m(0)))), & \gamma \geq 0, \\
\min(h^2, \max(0, -2\log((\gamma/m(0)))), & \gamma \leq 0.
\end{cases}
\]
Figure 1: $\rho_0(x)$ and $1 - \rho_0(x)$ versus $||x||$, as in Example 3.1, with $q = 2$, $\nu = .13$, $\omega = 2$, $\kappa = 4$. The value of $\nu$ corresponds to a prior choice $s = .5$.

Then

$$\rho_0(x) = \begin{cases} \tau + \frac{\gamma}{m(x)}, & 0 \leq ||x||^2 \leq d^2, \\ 1, & d^2 \leq ||x||^2 \leq h^2, \\ \tau + \frac{\gamma}{m(x)}, & 0 \leq ||x||^2 \leq d^2, \\ 0, & d^2 \leq ||x||^2 \leq h^2, \end{cases}$$

if $\gamma \geq 0$,

$$if \gamma \leq 0,$$

$$a = h^qV, \quad r = \frac{2^{-q/2}m(0)H(2h^2)}{c},$$

and

$$\xi_k = c^{-k} \int_{||x||^2 \leq d^2} \phi^k(x)dx = \begin{cases} \frac{d^qV}{H(d^2)}, & k = 0, \\ \frac{H(d^2)}{2^{-q/2}m(0)H(2h^2)}c, & k = 2. \end{cases}$$

Table 1 gives the minimising values of $\tau$ and $\gamma$ in some special cases with $c = .99$. We have taken $\omega = 2$ and $\kappa = 4$ in order to reflect the experimenter’s greater uncertainty about the validity of the model of the responses to the treatment. For $q = 1, 2, 3, 4$ we exhibit $a, s, d, \tau$ and $\gamma$ for a range of values of $\nu$. We give the percent relative (mse) efficiency $rme = 100 \times \mathcal{L}(\rho_K)/\mathcal{L}(\rho_0)$ enjoyed by $\rho_0$ over the naive choice $\rho_K(x)$, which minimises variance alone. We also give the corresponding relative (variance) efficiency $rve$ computed by using only the variance component of the loss. As one might expect the gains in $rme$ are greatest when $\nu$ is small and bias is of considerably greater concern than variance. This
is, or should be, the typical case. With respect to variance alone the minimax allocation schemes are relatively inefficient unless \( \nu \) is near 1. For \( q = 1 \) the comparable efficiencies of (14) with respect to \( \rho_K \), with \( \kappa(x) \equiv \kappa \) are given as well. See Figure 1 for plots of \( \rho_0(x) \) and \( 1 - \rho_0(x) \) versus \( ||x|| \), if \( q = 2 \) and \( \nu = .13 \), a value which ensures that the unconditional probability \( s = .5 \) of an assignment to the treatment is optimal.

Investigations with other values of \( \omega \) and \( \kappa \) lead to the same conclusions made evident by Table 1 and Figure 1. When there is relatively greater uncertainty about the model for the treatment than for the control, as expressed by \( \omega > 1 \) or by \( \kappa > 1 \), then with relatively high probability the treatment is to be assigned those subjects whose (standardised) covariates are larger in norm. The probability \( s \) is increasing in \( \omega \) and decreasing in \( \kappa \).

**APPENDIX: DERIVATIONS**

**Proof of Theorem 1:** We give the details of the derivation of MSE for Model 2, and then specialise the results to Model 1. First suppose that of the \( n \) subjects entering the study a total of \( n_i \), with covariates \( x_{ij}, j = 1, ..., n_i \), are assigned to group \( i \). Denote by \( y \) the \( n \times 1 \) data vector, with elements \( y_{ij} \) ordered lexicographically. Then

\[
y = V\psi + n^{-1/2}f + \Sigma^{1/2}\varepsilon,
\]

where the row of \( V \) corresponding to \( y_{ij} \) is \( v_{ij}^T := (u_i^T, z^T(x_{ij})) \), the corresponding element of \( f \) is \( f_i(x_{ij}) \), and the corresponding element of the diagonal matrix \( \Sigma^{1/2} \) is \( \sigma_i \).
The least squares estimate \( \hat{\psi} = \left( \frac{V^T V}{n} \right)^{-1} \frac{V^T y}{n} \) has covariance matrix

\[
\text{COV}[\hat{\psi}] = \frac{1}{n} \left( \frac{V^T V}{n} \right)^{-1} \left( \frac{V^T \Sigma V}{n} \right) \left( \frac{V^T V}{n} \right)^{-1}
\]

and bias vector

\[
E[\hat{\psi} - \psi] = \frac{1}{\sqrt{n}} \left( \frac{V^T V}{n} \right)^{-1} \frac{V^T f_n}{n}.
\]

Put \( P_i^n = n_i/n \). Then

\[
\frac{V^T V}{n} = \sum_{i=1}^p P_i^n \left( u_i \cdot \frac{1}{n_i} \sum_{j=1}^{n_i} z^T (x_{ij}) \right),
\]

\[
\frac{V^T \Sigma V}{n} = \sum_{i=1}^p P_i^n \sigma_i^2 \left( u_i \cdot \frac{1}{n_i} \sum_{j=1}^{n_i} z^T (x_{ij}) \right),
\]

\[
\frac{V^T f_n}{n} = \sum_{i=1}^p P_i^n \left( u_i \cdot \frac{1}{n_i} \sum_{j=1}^{n_i} f_i(x_{ij}) \right).
\]

By the Strong Law of Large Numbers these expressions of the form \( P_i^n \sum_{j=1}^{n_i} \Delta(x_{ij})/n_i \) tend, with probability 1 as \( n_i \to \infty \), to

\[
P(\text{group } i) \cdot E[\Delta(x)|i] = \int_S \Delta(x)m_i(x)dx
\]

and so, with probability 1, \( \frac{V^T V}{n} \to B \), \( \frac{V^T \Sigma V}{n} \to Q \) and \( \frac{V^T f_n}{n} \to b = b(f_1, ..., f_p) \) as defined at (5). Thus

\[
n\text{MSE}(\hat{\psi}) = n\text{COV}[\hat{\psi}] + \left( \sqrt{n}E[\hat{\psi} - \psi] \right) \left( \sqrt{n}E[\hat{\psi} - \psi] \right)^T \sim |B^{-1} (Q + bb^T) B^{-1}|
\]

and \( W_0 \hat{\theta} = W\hat{\psi} \) satisfies

\[
|n\text{MSE}(W_0 \hat{\theta})| \overset{a.s.}{\to} |W B^{-1} Q B^{-1} W^T| \left( 1 + b^T B^{-1} W^T (W B^{-1} Q B^{-1} W^T)^{-1} W B^{-1} b \right)
\]

\[
= |W B^{-1} Q B^{-1} W^T| (1 + ||Rb(f_1, ..., f_p)||^2).
\]

This gives (7). These expressions are valid for Model 1 as well, with the sole change \( v_i(x) := u_i \).

To maximise

\[
||Rb(f_1, ..., f_p)||^2 = \left| \sum_{i=1}^p \int_S Rv_i(x)m_i(x)f_i(x)dx \right|^2
\]
over \( f_1, \ldots, f_p \), we proceed by the method of Lagrange multipliers. A more formal derivation follows along the lines of the proof of Theorem 1 of Wiens (1992).

Let \( f_{i,0} \) and \( f_{i,1} \) satisfy (3) and (4) and set \( f_{i,t} = (1-t)f_{i,0} + tf_{i,1} \) for \( t \in [0,1] \). For Lagrange multipliers \( c_i \) and \( d_i \), consider

\[
\varphi(t_1, \ldots, t_p) := \left| \sum_{i=1}^{p} \int_{S} R v_i(x) m_i(x) f_{i,t}(x) \, dx \right|^2 + 2 \sum_{i=1}^{p} \int_{S} c_i^T v_i(x) f_{i,t}(x) \, dx - \sum_{i=1}^{p} d_i \int_{S} f_{i,t}^2(x) \, dx.
\]

We seek \( f_{1,0}, \ldots, f_{p,0} \) which satisfy the side conditions, and are such that \( \varphi \) is maximised at \( 0 \). The first order conditions \( \partial \varphi / \partial t_i |_{t_1 = \ldots = t_p = 0} = 0 \) become, with \( b_0 := b(f_{1,0}, \ldots, f_{p,0}) \),

\[
2 \int_{S} (f_{i,1}(x) - f_{i,0}(x)) \left\{ (R b_0)^T R v_i(x) m_i(x) + c_i^T v_i(x) - d_i f_{i,0}(x) \right\} \, dx, \quad i = 1, \ldots, p,
\]

for all \( f_{i,1} \), so that \( f_{i,0} \) is of the form

\[
f_{i,0}(x) = (m_i(x) \beta_i + \gamma_i)^T v_i(x) \quad (A.1)
\]

for vectors \( \beta_i \) and \( \gamma_i \) chosen to maximise \( ||R b_0||^2 \) and to satisfy the constraints.

Condition (3) gives \( \gamma_i = -A_i^+ B_i \beta_i \), so that

\[
f_{i,0}(x) = \beta_i^T (m_i(x) I_p - B_i A_i^+) v_i(x).
\]

Condition (4) becomes

\[
\beta_i^T G_i \beta_i \leq \eta_i^2, \quad i = 1, \ldots, p \quad (A.2)
\]

and \( ||R b_0||^2 = ||R \sum_{i=1}^{p} G_i \beta_i||^2 \) is to be maximised subject to (A.2).

Note that

\[
G_i = \int_{S} (m_i(x) I_p - B_i A_i^+) v_i(x) v_i^T(x) (m_i(x) I_p - A_i^+ B_i) \, dx,
\]

so that \( G_i \) is positive semi-definite. With \( \alpha_i := G_i^{1/2} \beta_i / \eta_i \), we equivalently maximize \( ||R \sum_{i=1}^{p} \eta_i G_i^{1/2} \alpha_i||^2 \) over \( \alpha_1, \ldots, \alpha_p \) of norm not exceeding 1. For this, we note that

\[
||R \sum_{i=1}^{p} \eta_i G_i^{1/2} \alpha_i||^2 = \left( \max_{||\alpha||=1} |\alpha^T R \sum_{i=1}^{p} \eta_i G_i^{1/2} \alpha_i| \right)^2.
\]
so that
\[
\max_{||\alpha_1||,...,||\alpha_p||\leq 1} ||R \sum_{i=1}^{p} \eta_i G_{i}^{1/2} \alpha_i||^2 = \left( \max_{||\alpha||=1} \max_{||\alpha_1||,...,||\alpha_p||\leq 1} |\alpha^T R \sum_{i=1}^{p} \eta_i G_{i}^{1/2} \alpha_i| \right)^2.
\]
The maxima over the $\alpha_i$ are attained at $G_{i}^{1/2} R^T \alpha / ||G_{i}^{1/2} R^T \alpha||$, yielding (8). \(\square\)

The proof of Theorem 2 uses the following result.

**Lemma A1.** If the rows of $W_{p-1 \times p}$ are mutually orthonormal contrasts, so that $Q := \left( \frac{1}{\sqrt{p}} \begin{bmatrix} W^T \end{bmatrix} \right)^T$ is an orthogonal matrix, and if $D_{p \times p}$ is a non-singular matrix, then

(i) $(WD^{-1} W^T)^{-1} = WDW^T - \frac{WD11^T DW^T}{1^T D1}$,

(ii) $W^T (WD^{-1} W^T)^{-1} W = D - \frac{D11^T D}{1^T D1}$,

(iii) $|WD^{-1} W^T| = \frac{1^T D1}{p|D|}$.

**Proof:** For (i), note that $QD^{-1} Q^T = (QDQ^T)^{-1}$. The lower right hand block of $QD^{-1} Q^T$ is $WD^{-1} W^T$; that of $(QDQ^T)^{-1}$ is the inverse of the matrix on the right hand side of (i). Now (ii) is a direct calculation. Statement (iii) is obtained by writing $QD^{-1} Q^T$ in partitioned form and substituting (ii) into

\[
|D^{-1}| = |QD^{-1} Q^T| = \frac{|WD^{-1} W^T|}{p} \left\{ 1^T D^{-1} - 1^{-1} D^{-1} W^T (WD^{-1} W^T)^{-1} W D^{-1} \right\}.
\]

\(\square\)

**Proof of Theorem 2:** First consider Model 1. We calculate that $B = diag(s_1, \ldots, s_p)$, $B^{-1} Q B^{-1} = \sigma_1^2 D^{-1}$ and that $G_i = g_i u_i u_i^T$. Now (10) follows, upon using Lemma A1 (iii).

An upper bound on the term in braces in (10) is

\[
\sum_{i=1}^{p} \eta_i \sqrt{g_i} \max_{||\alpha||=1} |\alpha^T R u_i| = \sum_{i=1}^{p} \eta_i \sqrt{g_i} ||R u_i|| = \sum_{i=1}^{p} \eta_i \sqrt{g_i (R^T R)_{ii}}.
\]

(A.3)
The bound is exact if $p = 2$, since then $\alpha = \pm 1$ and there is nothing to maximise.

By Lemma A1 (ii),

\[
R^T R = \sigma_1^{-2} B^{-1} \left( D - \frac{D11^T D}{tr D} \right) B^{-1}.
\]

(A.4)
Substituting (A.4) into (A.3) and simplifying gives (11).

For Model 2, note that the restrictions we have imposed imply that

\[
\int_S z(x) m_{ik}^k(x) dx = 0, \ k = 0, 1, 2, \ i = 1, \ldots, p.
\]
It follows that the matrices $B$, $Q$, and $G_i$ are all block-diagonal. In each case the first of the two blocks is the corresponding $p \times p$ matrix arising in Model 1. The second $q \times q$ block depends on the covariates. Upon evaluating (7) and (8) with $W = \begin{pmatrix} W_0 & 0 \end{pmatrix}$ however, these lower blocks are eliminated. The result is that (10) holds in both models. □

**Proof of Lemma 1:**

$$\varphi''(t) = \frac{\left(\int_S (l_1(x) - l_0(x))^2 \, dx\right) \left(\int_S l_t^2(x) \, dx\right) - \left(\int_S (l_1(x) - l_0(x)) l_t(x) \, dx\right)^2}{\left(\int_S l_t^2(x) \, dx\right)^{3/2}}$$

is non-negative, by the Cauchy-Schwarz inequality. □

**REFERENCES**


