20

Robustness of Design

Douglas P. Wiens

CONTENTS

20.1 Introduction

Robustness of design is a vast topic – the Current Index to Statistics lists over 700 articles purporting to discuss it – and so we shall first clarify our restricted interpretation of the term. By *robustness of design* we shall mean the scenario outlined in the next few paragraphs.

An investigator anticipates planning a study that will result in a number of observations on a random variable y, whose probability distribution – often merely through its expected value – depends on a vector $\boldsymbol{x}$ of covariates *that can be set by the investigator*, hence the design. After the data are gathered, the relationship between y and $\boldsymbol{x}$ is to be assessed. This will generally involve both estimation and prediction and is often done in the context of a particular model *of which the experimenter might have only partial knowledge and in which he might have little faith* – hence the robustness requirement.

Robustness has numerous meanings in statistics. The notion appears to have been introduced by Box (1953) and was given a firm mathematical basis by Huber (1964, 1981), for whom it generally – but certainly not exclusively – meant the relative insensitivity of a statistical procedure to departures from the assumed Gaussian error distribution. In design, the usual performance measures depend on the error distribution only through the first two moments, and beyond this, the distributional shape is not so relevant. One does however have in mind a particular model to be fitted once the data are gathered. In *classical* optimal design theory, one believes explicitly that the model one fits is the correct one and measures the quality of a design through a *loss function* such as the determinant, or trace, or maximum eigenvalue of the covariance matrix, corresponding to the well-known D, A and E-optimality criteria discussed in Chapters 1 and 2. In *model robust* design theory, one instead anticipates that the model that will be fitted by the experimenter is not necessarily the true one – a simple example to bear in mind is that of fitting a straight line regression when the true response function is possibly not exactly linear in the covariate – and so the loss function highlights some more general feature such as the mean squared error (*mse*). This will of course depend on the true, rather than fitted, model, and so one seeks a design minimizing some scalar quantity summarizing the increased loss – perhaps the maximum, or average, of the *mse* over the predicted values – as the true model varies.

It is the notion of model robustness on which this chapter is primarily focussed. An experimenter plans to fit a particular model to his data while realizing that any model is at best an approximation. He seeks protection, at the design stage, from increased loss incurred by model misspecification.

Some other robustness concepts – certainly worthy of discussion, but which space limitations prevent us from discussing more fully here – are as discussed as follows:

- *Criterion robustness*: Here, one aims to find a design optimizing a mixture of criteria, for instance, discriminatory power to identify a true model among a class of competitors, and estimation efficiency at this model. In some cases – for instance, if the competing models are all only vaguely specified – this can be viewed as a sub-topic of model robustness and will be discussed in such a context. For other notions of criterion robustness – for instance, designing to minimize some mixture of the D, A and E-optimality measures applied to the covariance matrix when the model is not in doubt – a pioneering reference is Kiefer (1975).
- *Robustness of designs for comparative experiments*: The robustness issues, and techniques for handling them, are very different for block designs, factorial designs, etc. than for regression-based experiments. The robustness issues tend to revolve around effects whose estimation is precluded by the design – because of blocking, or aliasing, for instance – and the concentration is more on assessing the robustness of existing designs, and choosing among them, than on constructing optimally robust designs.

In the next two sections of this chapter, we discuss model robustness in some generality, classifying by the types of model departures of particular interest. Such a classification is quite crude, and so in Section 20.4 we look at a number of special applications in which robustness is sought. Several, evidently open, problems of potential interest are noted. Our coverage of the field is of necessity quite selective, and we apologize to those whose contributions we have had to omit.

20.2 Robustness against a Misspecified Response Function

In planning a design strategy that is to be robust against response functions other than those which will be fitted by the analyst, one must first characterize the class of alternatives. The seminal work here is by Box and Draper (1959), who considered, among others, the problem of designing for a polynomial response when the class of alternatives consists of all such functions of a given higher degree. As a simple yet motivating example of their results, suppose that one is to fit a straight line in an independent variable $x \in [-1, 1]$ when in fact the mean response is $E[y(x)] = \phi_0 + \phi_1 x + \phi_2 x^2$ for parameters $\boldsymbol{\phi}$. Assume uncorrelated errors with common variance σ_ε^2. With (not necessarily distinct) design points $x_1, \ldots, x_n$, and with $\boldsymbol{X}$ denoting the model matrix with the ith row $\boldsymbol{f}'(x_i) = (1, x_i)$, the model to be fitted to the vector $\boldsymbol{y}$ of observations is $E[\boldsymbol{y}] = \boldsymbol{X}\theta$ for some $\theta = (\theta_0, \theta_1)'$; neither θ_0 nor θ_1 need equal the corresponding parameter in the ϕ-parameterization describing the true model. The least squares estimate (*lse*) is $\hat{\theta} = \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{y}$. Define $\tau_k = \sum x_i^k/n$ and assume that $\tau_1 = \tau_3 = 0$, as, for instance, is the case if the design is symmetric. Then one finds that, under the true quadratic model, the mean vector and covariance matrix of $\hat{\theta}$ are $(\phi_0 + \tau_2\phi_2, \phi_1)'$ and $\left(\sigma_\varepsilon^2/n\right) diag\left(1, \tau_2^{-1}\right)$, respectively, so that the predictions $\hat{y}(x) = (1, x)\,\hat{\theta}$ have *mse*

$$\text{MSE}\left[\hat{y}(x)\right] = E\left[\left\{\hat{y}(x) - E[y(x)]\right\}^2\right] = \frac{\sigma_\varepsilon^2}{n}\left(1 + \frac{x^2}{\tau_2}\right) + \left(\phi_2\left(\tau_2 - x^2\right)\right)^2.$$

A common measure of performance is the integrated mean squared error (*imse*) of the predictors, which in this instance is

$$\text{IMSE} = \int_{-1}^{1} \text{MSE}\left[\hat{y}(x)\right] dx = \left\{\frac{2\sigma_\varepsilon^2}{n}\left(1 + \frac{1}{3\tau_2}\right)\right\} + \left\{2\phi_2^2\left(\left(\tau_2 - \frac{1}{3}\right)^2 + \frac{4}{45}\right)\right\}. \tag{20.1}$$

The first term in braces in (20.1) is the integrated variance and is minimized by the design – simultaneously optimal with respect to several criteria (*D*-, *A*-, *E*-, etc.) – with half of the observations made at each of $x = \pm 1$. The second – the integrated bias, dominating the first once n is sufficiently large – is minimized if $\tau_2 = 1/3$. This can be attained in many ways, but we note that it is in particular the second moment of the (continuous) uniform distribution on $[-1, 1]$. If higher order alternatives are considered, then minimization of this second term requires correspondingly higher order moments to agree with those of the uniform distribution. A design approximating the uniform is the equally spaced design $x_i = -1 + 2(i-1)/(n-1)$, with $\tau_2 = (1/3) + 2/(3(n-1))$. The conclusion that can be drawn here and that was drawn in the other cases considered is that "... the optimal design in typical situations in which both variance and bias occur is very nearly the same as would be obtained if *variance were ignored completely* and the experiment designed so as to *minimize the bias alone*" (Box and Draper 1959, p. 622).

To generalize the setup of this example and to set notation to be used throughout this chapter, suppose that one has a p-vector $\boldsymbol{f}(\boldsymbol{x})$ of regressors, each element of which is a

function of q functionally independent variables $\boldsymbol{x} = \left(x_1, \ldots, x_q\right)'$, with $\boldsymbol{x}$ to be chosen from a *design space* $\mathcal{X}$. Then the fitted model is $E\left[y\left(\boldsymbol{x}\right)\right] = \boldsymbol{f}'\left(\boldsymbol{x}\right)\boldsymbol{\theta}$, with alternatives

$$E\left[y\left(\boldsymbol{x}\right)\right] = \boldsymbol{f}'\left(\boldsymbol{x}\right)\boldsymbol{\theta} + \psi\left(\boldsymbol{x}\right), \tag{20.2}$$

for some function ψ. There is an immediate problem concerning the interpretation of $\boldsymbol{\theta}$, since in the alternate models, one might equally well write $E\left[y\left(\boldsymbol{x}\right)\right] = \boldsymbol{f}'\left(\boldsymbol{x}\right)\left(\boldsymbol{\theta} + \boldsymbol{\phi}\right) + \left(\psi\left(\boldsymbol{x}\right) - \boldsymbol{f}'\left(\boldsymbol{x}\right)\boldsymbol{\phi}\right)$ for arbitrary $\boldsymbol{\phi}$, whence the parameter vector is not identifiable. This is avoided by first *defining* the target parameter, for instance, by

$$\boldsymbol{\theta} = \arg\min_{\boldsymbol{\eta}} \int_{\mathcal{X}} \left(E\left[y\left(\boldsymbol{x}\right)\right] - \boldsymbol{f}'\left(\boldsymbol{x}\right)\boldsymbol{\eta}\right)^2 d\boldsymbol{x}, \tag{20.3}$$

and then *defining*

$$\psi\left(\boldsymbol{x}\right) = E\left[y\left(\boldsymbol{x}\right)\right] - \boldsymbol{f}'\left(\boldsymbol{x}\right)\boldsymbol{\theta}; \tag{20.4}$$

this leads to the orthogonality requirement

$$\int_{\mathcal{X}} \boldsymbol{f}\left(\boldsymbol{x}\right)\psi\left(\boldsymbol{x}\right) d\boldsymbol{x} = \boldsymbol{0}. \tag{20.5}$$

Under the very mild assumption that the matrix $\boldsymbol{A} = \int_{\mathcal{X}} \boldsymbol{f}\left(\boldsymbol{x}\right)\boldsymbol{f}'\left(\boldsymbol{x}\right) d\boldsymbol{x}$ be invertible – equivalent to the statement that if $\boldsymbol{c}'\boldsymbol{f}\left(\boldsymbol{x}\right) = 0$ (a.e. $\boldsymbol{x} \in \mathcal{X}$), then $\boldsymbol{c} = \boldsymbol{0}$ – the parameter defined by (20.4) and (20.5) is unique:

$$\boldsymbol{\theta} = \boldsymbol{A}^{-1} \int_{\mathcal{X}} \boldsymbol{f}\left(\boldsymbol{x}\right) E\left[y\left(\boldsymbol{x}\right)\right] d\boldsymbol{x}.$$

In the motivating example given earlier, $\boldsymbol{f}\left(x\right) = \left(1, x\right)'$ and (20.5) applied to quadratic alternatives forces $\psi\left(x\right) = \phi_2\left(1/3 - x^2\right)$ for some ϕ_2.

In what follows, we identify a design, denoted ξ, with its design measure – a probability measure $\xi\left(d\boldsymbol{x}\right)$ on $\mathcal{X}$. If n_i of the n observations are to be made at $\boldsymbol{x}_i$, we also write $\xi_i = \xi\left(\boldsymbol{x}_i\right) = n_i/n$. Define

$$\boldsymbol{M}_\xi = \int_{\mathcal{X}} \boldsymbol{f}\left(\boldsymbol{x}\right)\boldsymbol{f}'\left(\boldsymbol{x}\right)\xi\left(d\boldsymbol{x}\right), \tag{20.6a}$$

$$\boldsymbol{b}_{\psi,\xi} = \int_{\mathcal{X}} \boldsymbol{f}\left(\boldsymbol{x}\right)\psi\left(\boldsymbol{x}\right)\xi\left(d\boldsymbol{x}\right), \tag{20.6b}$$

and assume that $\boldsymbol{M}_\xi$ is invertible. The covariance matrix of the *lse* $\hat{\boldsymbol{\theta}}$ is $\left(\sigma_\varepsilon^2/n\right)\boldsymbol{M}_\xi^{-1}$, the bias is $E\left[\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\right] = \boldsymbol{M}_\xi^{-1}\boldsymbol{b}_{\psi,\xi}$, and the general version of (20.1) is found to be

$$\text{IMSE} = \int_{\mathcal{X}} \text{MSE}\left[\hat{y}\left(\boldsymbol{x}\right)\right] d\boldsymbol{x} = \frac{\sigma_\varepsilon^2}{n}\text{trace}\left(\boldsymbol{A}\boldsymbol{M}_\xi^{-1}\right) + \boldsymbol{b}'_{\psi,\xi}\boldsymbol{M}_\xi^{-1}\boldsymbol{A}\boldsymbol{M}_\xi^{-1}\boldsymbol{b}_{\psi,\xi} + \int_{\mathcal{X}} \psi^2\left(\boldsymbol{x}\right) d\boldsymbol{x}. \tag{20.7}$$

It is obvious from (20.7), and true even if *imse* is not the loss function, that one must bound the influence of $\psi(\cdot)$, to complete the definition of a class Ψ of alternatives in (20.2). For straight line regression, Huber (1975) defines Ψ by (20.5) together with

$$\int_{\mathcal{X}} \psi^2(\boldsymbol{x})\, d\boldsymbol{x} \leq \tau^2/n, \tag{20.8}$$

for a given constant τ; this class was generalized to other scenarios by Wiens (1992). That the bound be $O\left(n^{-1}\right)$ is required for a sensible asymptotic treatment based on the *mse* – it forces the bias of the estimates to decrease at the same rate as their standard error. Marcus and Sacks (1976), Pesotchinsky (1982) and Li and Notz (1982) instead take

$$|\psi(\boldsymbol{x})| \leq \phi(\boldsymbol{x}), \tag{20.9}$$

for a specified function $\phi(\cdot)$. The resulting optimal *minimax* designs (maximize over ψ and then minimize over ξ) depend on the form of ϕ, but commonly – as in the three articles just mentioned – the design mass is concentrated on a small number of extreme points of $\mathcal{X}$. This precludes an investigation of the response function in the interior of $\mathcal{X}$ and so is clearly not robust.

The class Ψ defined by (20.5) and (20.8) is not immune from criticism – it is so rich that any design with finite maximum loss is necessarily absolutely continuous, hence must be approximated in order to be implemented. That this is so is intuitively clear: the Lebesgue integrals defining Ψ may be modified on sets of Lebesgue measure zero; thus, if ξ places mass on any such set – as does any discrete measure – then one can choose ψ to be arbitrarily large there, thus exploding the elements of $\boldsymbol{b}_{\psi,\xi}$. A formal proof may be found in Heo et al. (2001). However, as in Wiens (1992), "Our attitude is that an approximation to a design which is robust against more realistic alternatives is preferable to an exact solution in a neighbourhood which is unrealistically sparse." To implement a continuous design on an interval, one might place the design points at the quantiles: $x_i = \xi^{-1}((i-0.5)/n)$, $i = 1, \ldots, n$; this empirical approximation is optimal in a minimum Kolmogorov discrepancy sense (Fang and Wang 1994, §1.2) and approaches ξ weakly as $n \to \infty$. For this case, and especially for multidimensional designs, there is a variety of such approximation methods (see, e.g., Xu and Yuen 2011). Rounding strategies for implementing discrete designs with continuous weights ξ_i are discussed by Pukelsheim and Rieder (1992).

To obtain a design, robust with respect to (20.5) and (20.8), one first maximizes (20.7) under these constraints. In contrast to the next step – minimization over the class of designs – this can be done in complete generality. Let $m(\boldsymbol{x})$ be the density of ξ and define $\boldsymbol{H}_\xi = \boldsymbol{M}_\xi \boldsymbol{A}^{-1} \boldsymbol{M}_\xi$, $\boldsymbol{K}_\xi = \int_{\mathcal{X}} \boldsymbol{f}(\boldsymbol{x}) \boldsymbol{f}'(\boldsymbol{x})\, m^2(\boldsymbol{x})\, d\boldsymbol{x}$ and

$$\boldsymbol{G}_\xi = \boldsymbol{K}_\xi - \boldsymbol{H}_\xi = \int_{\mathcal{X}} \left[\left(m(\boldsymbol{x})\, \boldsymbol{I}_p - \boldsymbol{M}_\xi \boldsymbol{A}^{-1}\right) \boldsymbol{f}(\boldsymbol{x})\right] \left[\left(m(\boldsymbol{x})\, \boldsymbol{I}_p - \boldsymbol{M}_\xi \boldsymbol{A}^{-1}\right) \boldsymbol{f}(\boldsymbol{x})\right]' d\boldsymbol{x}.$$

The matrix $\boldsymbol{G}_\xi$ is clearly positive semidefinite; assume for the moment that it is positive definite and define a function $\mathbf{r}(\boldsymbol{x}) = \left(\tau/\sqrt{n}\right) \boldsymbol{G}_\xi^{-1/2} \left(m(\boldsymbol{x})\, \boldsymbol{I}_p - \boldsymbol{M}_\xi \boldsymbol{A}^{-1}\right) \boldsymbol{f}(\boldsymbol{x})$. We have the following identities:

1. $\int_{\mathcal{X}} \boldsymbol{r}(\boldsymbol{x}) \boldsymbol{r}'(\boldsymbol{x})\, d\boldsymbol{x} = \left(\tau^2/n\right) \boldsymbol{I}_p$.
2. $\int_{\mathcal{X}} \boldsymbol{f}(\boldsymbol{x}) \boldsymbol{r}'(\boldsymbol{x})\, d\boldsymbol{x} = \mathbf{0}_{p \times p}$.

3. $\int_{\mathcal{X}} \boldsymbol{f}(\boldsymbol{x})\boldsymbol{r}'(\boldsymbol{x})m(\boldsymbol{x})\,d\boldsymbol{x} = \left(\tau/\sqrt{n}\right)\boldsymbol{G}_{\xi}^{1/2}$.
4. $\int_{\mathcal{X}} \boldsymbol{r}(\boldsymbol{x})\psi(\boldsymbol{x})\,d\boldsymbol{x} = \left(\tau/\sqrt{n}\right)\boldsymbol{G}_{\xi}^{-1/2}\boldsymbol{b}_{\psi,\xi}$.

As in Wiens (1992), it follows from (1) and (2) that the class $\Psi_0 = \left\{\psi_{\boldsymbol{\beta}}(\boldsymbol{x}) = \mathbf{r}'(\boldsymbol{x})\boldsymbol{\beta} \mid \|\boldsymbol{\beta}\| = 1\right\}$ is a subclass of Ψ (with equality in (20.8)) that is least favourable in that the supremum, over Ψ, of (20.7) is attained by a member of Ψ_0. To see this last point, let $\psi \in \Psi$ be arbitrary and set $\boldsymbol{\beta}_* = \boldsymbol{G}_{\xi}^{-1/2}\boldsymbol{b}_{\psi,\xi} \Big/ \left\|\boldsymbol{G}_{\xi}^{-1/2}\boldsymbol{b}_{\psi,\xi}\right\|$. By (3), (20.7) evaluated at $\psi_{\boldsymbol{\beta}_*}$ gives

$$\text{IMSE}_{|\psi_{\boldsymbol{\beta}_*}} = \frac{\sigma_{\varepsilon}^2}{n}\text{trace}\left(\boldsymbol{A}\boldsymbol{M}_{\xi}^{-1}\right) + \frac{\tau^2}{n}\frac{\boldsymbol{b}'_{\psi,\xi}\boldsymbol{H}_{\xi}{}^{-1}\boldsymbol{b}_{\psi,\xi}}{\left\|\boldsymbol{G}_{\xi}^{-1/2}\boldsymbol{b}_{\psi,\xi}\right\|^2} + \frac{\tau^2}{n}. \tag{20.10}$$

The Cauchy–Schwarz inequality, followed by (4), gives

$$\frac{\tau^2}{n} \geq \left|\int_{\mathcal{X}} \psi(\boldsymbol{x})\,\psi_{\boldsymbol{\beta}_*}(\boldsymbol{x})\,d\boldsymbol{x}\right| = \frac{\tau}{\sqrt{n}}\left\|\boldsymbol{G}_{\xi}^{-1/2}\boldsymbol{b}_{\psi,\xi}\right\|,$$

so that the *imse* (20.10) is at least as large as that – (20.7) – evaluated at ψ.

Evaluating (20.7) at $\psi_{\boldsymbol{\beta}}$ for arbitrary $\boldsymbol{\beta}$ gives

$$\text{IMSE}_{|\psi_{\boldsymbol{\beta}}} = \frac{\sigma_{\varepsilon}^2}{n}\text{trace}\left(\boldsymbol{A}\boldsymbol{M}_{\xi}^{-1}\right) + \frac{\tau^2}{n}\boldsymbol{\beta}'\left[\boldsymbol{G}_{\xi}^{1/2}\boldsymbol{H}_{\xi}^{-1}\boldsymbol{G}_{\xi}^{1/2} + \boldsymbol{I}_p\right]\boldsymbol{\beta};$$

now maximizing over $\boldsymbol{\beta}$ yields the result that $\max_{\Psi}\text{IMSE}$ is $\left(\sigma_{\varepsilon}^2 + \tau^2\right)/n$ times

$$\mathcal{L}_{\nu}(\xi) = (1-\nu)\,\text{trace}\left(\boldsymbol{A}\boldsymbol{M}_{\xi}^{-1}\right) + \nu\text{ch}_{\max}\left(\boldsymbol{K}_{\xi}\boldsymbol{H}_{\xi}^{-1}\right), \tag{20.11}$$

where $\nu = \tau^2/\left(\sigma_{\varepsilon}^2 + \tau^2\right)$ and $\text{ch}_{\max}$ denotes the maximum eigenvalue. If $\boldsymbol{G}$ is singular, then one first perturbs it to make it nonsingular and then passes to the limit – details in Heo et al. (2001).

One is now to choose $\nu \in [0,1]$, reflecting the relative importance to the experimenter of errors due to bias rather than to variance, and minimize $\mathcal{L}_{\nu}(\xi)$. This step is highly dependent on the form of the model being fitted. In some simple cases, it can be done analytically, using variational methods to minimize $\mathcal{L}_{\nu}(\xi)$ over $\xi' = m$ subject to various side conditions. For instance, in straight line regression – under the restriction that the design be symmetric – the maximization can be carried out in two stages. At the first stage, one imposes the conditions that m has an integral of one and a fixed second moment $\gamma = \int_{\mathcal{X}} x^2 m(x)dx$; this fixes $\boldsymbol{M}_{\xi}$ and hence the first term in (20.11), and one then seeks a nonnegative, symmetric function minimizing the second term subject to these side conditions. The result is a partially minimized value $\min \mathcal{L}_{\nu}(\xi|\gamma)$; at the second stage, a minimization over γ is performed. See Huber (1981) for details. For bivariate regression (without interactions) on a spherical design space, details are in Wiens (1990) – see Figure 20.1 for a plot of $m(\boldsymbol{x})$, minimizing $\mathcal{L}_{\nu}(\xi)$ when $\nu = 0.36$. We note that in both of these cases, a proof that the optimal design is *necessarily* symmetric is still outstanding.

In these and the other examples in Wiens (1990, 1992), a difficulty encountered is that it is not clear which of the eigenvalues $e_1(\xi), \ldots, e_p(\xi)$ of $\boldsymbol{K}_{\xi}\boldsymbol{H}_{\xi}^{-1}$ – note that these depend

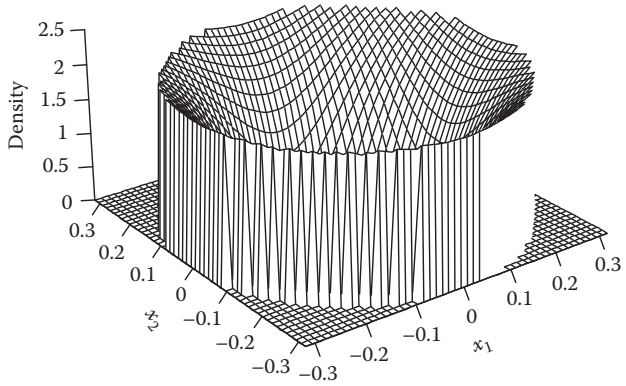

FIGURE 20.1
Minimax design density for bivariate regression without interaction on a spherical design space.

on the design – will turn out to be the maximum one when evaluated at the final, minimizing design. The usual approach is to attempt to guess correctly which one is to be optimized and to then verify at the end that the guess was correct. There are cases however in which this approach fails. One of the simplest of these is quadratic regression ($p = 3$), where for a large range of values of ν, it happens that each design ξ_k derived to minimize $l_k(\xi) = (1-\nu)\text{trace}\left(\boldsymbol{A}\boldsymbol{M}_{\xi}^{-1}\right) + \nu e_k(\xi)$ turns out to result in a maximum eigenvalue in (20.11) that is not the one that was minimized: $\text{ch}_{\max}\left(\boldsymbol{K}_{\xi_k}\boldsymbol{H}_{\xi_k}^{-1}\right) \neq e_k(\xi_k)$. In such cases, a possible approach, detailed in Daemi and Wiens (2012), is to find designs ξ_k minimizing the kth of the competing forms $\{l_1(\xi), \ldots, l_p(\xi)\}$ of $\mathcal{L}_\nu(\xi)$ subject to the side condition that $e_k(\xi_k)$ exceed $e_i(\xi_k)$ for $i \neq k$; the optimal design is that for which $l_k(\xi_k) = \min\{l_1(\xi_1), \ldots, l_p(\xi_p)\}$.

Shi et al. (2003) instead minimize the nondifferentiable functional $\mathcal{L}_\nu(\xi)$ using nonsmooth optimization methods. Another approach is to restrict to a smaller but more tractable class of designs. Heo et al. (2001) consider the class of designs with densities of the form $m_{\boldsymbol{\beta}}(\boldsymbol{x}) = \max\left(0, \sum_j \beta_j f_j\left(x_1^2, \ldots, x_q^2\right)\right)$, where $\boldsymbol{f}(\boldsymbol{x}) = \left(f_1(\boldsymbol{x}), \ldots, f_p(\boldsymbol{x})\right)'$ in (20.2), and minimize (20.11) numerically over $\boldsymbol{\beta}$ subject to the requirement that the arguments of m be exchangeable. Design points are then chosen in such a way that the resulting design has empirical moments matching those of the optimal $m_{\boldsymbol{\beta}}(\boldsymbol{x})$, to as high an order as possible. For fitting a full second-order model, including linear, quadratic and interaction terms, in two variables x_1 and x_2 with $n = 48$ and $\nu = 1/6$, this yields the design in Figure 20.2.

The situation is somewhat simpler, with little loss of generality, in a finite design space. In practice, one is very often restricted to choosing from a finite, if large, set of levels of the independent variables. Li (1984) proposed designs for straight line regression, robust against departures satisfying (20.5) and (20.9) with constant ϕ, concentrated on sets of equally spaced points in the interval of interest. The resulting minimax designs spread out their mass near the end points of this interval. This is a recurring theme – a naive yet sensible and *near optimal* method of robustifying a design is to take the replicates in the classically optimal design for the model in question and spread these out into nearby clusters. This paradigm is also exemplified in Figure 20.2.

Fang and Wiens (2000) continue the approach of Li (1984). They take an N-element design space $\mathcal{X}$ and discretize the definition of Ψ by replacing the integrals in (20.3),

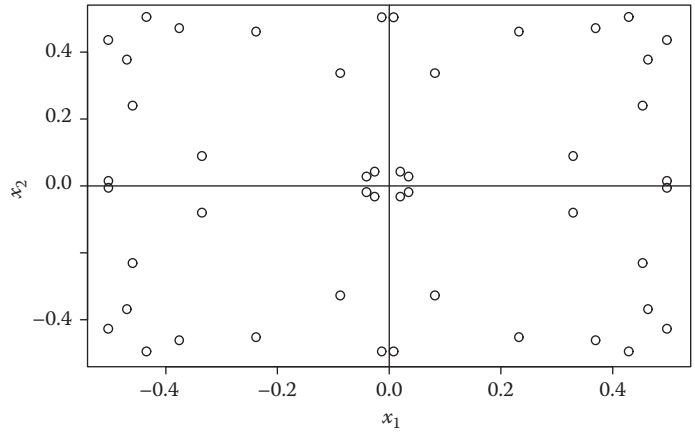

FIGURE 20.2
Restricted minimax design for fitting a full second-order model. (From Heo, G. et al., *Can. J. Stat.*, 29, 117, 2001.)

(20.5) and (20.8) by sums over $\mathcal{X}$. To maximize the *imse* $N^{-1}\sum_{i=1}^{N} E\left[\left\{\hat{y}\left(\boldsymbol{x}_i\right) - E\left[y\left(\boldsymbol{x}_i\right)\right]\right\}^2\right]$, it is convenient to temporarily transform to an orthogonal basis. Let $\boldsymbol{Q}_1$ be an $N \times p$ matrix whose columns form an orthogonal basis for the column space of the matrix with rows $\left\{\boldsymbol{f}'\left(\boldsymbol{x}\right) \mid \boldsymbol{x} \in \mathcal{X}\right\}$; this is computed in the QR decomposition of this matrix. Augment $\boldsymbol{Q}_1$ by $\boldsymbol{Q}_2 : N \times (N-p)$ whose columns form an orthogonal basis for the orthogonal complement of this space. Then $\left(\boldsymbol{Q}_1 \vdots \boldsymbol{Q}_2\right)$ is an orthogonal matrix and $\boldsymbol{\psi} = \left(\psi\left(\boldsymbol{x}_1\right), \ldots, \psi\left(\boldsymbol{x}_N\right)\right)'$ is necessarily of the form $\boldsymbol{\psi} = \left(\tau/\sqrt{n}\right)\boldsymbol{Q}_2\boldsymbol{c}$, where $\|\boldsymbol{c}\| \leq 1$. Define $\tilde{\boldsymbol{A}} = N^{-1}\sum_{i=1}^{N}\boldsymbol{f}\left(\boldsymbol{x}_i\right)\boldsymbol{f}'\left(\boldsymbol{x}_i\right)$, $\tilde{\boldsymbol{M}}_\xi = \sum_{i=1}^{N}\xi_i\boldsymbol{f}\left(\boldsymbol{x}_i\right)\boldsymbol{f}'\left(\boldsymbol{x}_i\right)$ and $\boldsymbol{D}_\xi = diag\left(\xi_1, \ldots, \xi_N\right)$. Then the analogue of (20.7) is

$$\text{IMSE} = \frac{\sigma_\varepsilon^2}{n}\text{trace}\left(\tilde{\boldsymbol{A}}\tilde{\boldsymbol{M}}_\xi^{-1}\right) + \frac{\tau^2}{nN}\boldsymbol{c}'\left[\boldsymbol{Q}_2'\boldsymbol{D}_\xi\boldsymbol{Q}_1\left(\boldsymbol{Q}_1'\boldsymbol{D}_\xi\boldsymbol{Q}_1\right)^{-2}\boldsymbol{Q}_1'\boldsymbol{D}_\xi\boldsymbol{Q}_2 + \boldsymbol{I}_{N-p}\right]\boldsymbol{c}. \tag{20.12}$$

Carrying out the maximization over $\boldsymbol{c}$ and returning to the original notation gives a direct analogue of (20.11): in terms of $\tilde{\boldsymbol{K}}_\xi = \sum_{i=1}^{N}\xi_i^2\boldsymbol{f}\left(\boldsymbol{x}_i\right)\boldsymbol{f}'\left(\boldsymbol{x}_i\right)$ and $\tilde{\boldsymbol{H}}_\xi = \tilde{\boldsymbol{M}}_\xi\tilde{\boldsymbol{A}}^{-1}\tilde{\boldsymbol{M}}_\xi$, the maximum loss is $\left(\sigma_\varepsilon^2 + \tau^2\right)/n$ times

$$\tilde{\mathcal{L}}_\nu\left(\xi\right) = (1-\nu)\,\text{trace}\left(\tilde{\boldsymbol{A}}\tilde{\boldsymbol{M}}_\xi^{-1}\right) + \nu\text{ch}_{\max}\left(\tilde{\boldsymbol{K}}_\xi\tilde{\boldsymbol{H}}_\xi{}^{-1}\right). \tag{20.13}$$

The minimization is carried out numerically – by simulated annealing, or via a genetic algorithm – yielding *exact*, that is, integer-valued, designs. See Figure 20.3 for an example, with $N = 40$, $n = 20$ and $\nu = .5$; this again illustrates the aforementioned paradigm.

This summary of possible approaches is by no means exhaustive. For instance, the use of *imse* to measure loss, while attractive, is not universal – one can instead apply the usual *D*-, *G*-, *A*- or *E*-criteria to the *mse* matrix of the regression parameters, rather than to the covariance matrix alone. Pesotchinsky (1978, 1982) measures the loss via a general L^p norm of the eigenvalues of the *mse* matrix; Marcus and Sacks (1976) and Li and Notz (1982) use a weighted trace of this matrix. Wiens (1993) maximizes the coverage probability of confidence ellipsoids. Zhou (2008) assumes a finite design space and obtains minimax *D*-optimal

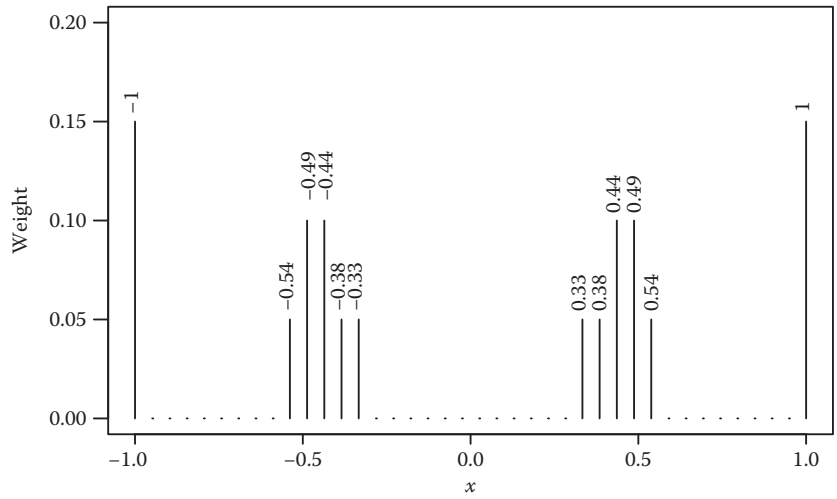

FIGURE 20.3
Minimax design ($n = 20$) for approximate cubic regression when bias and variance receive equal weight.

designs, with the maximum taken over the discrete version of Ψ described earlier. Adewale and Wiens (2006) *average* the *imse* (rather than maximizing it) with respect to a uniform prior on $\boldsymbol{c}$ in (20.12). Exact designs are then obtained by simulated annealing.

Various approaches, besides those described thus far, have been proposed with the aim of restricting Ψ while still hoping to capture a broad spectrum of alternatives. Fedorov et al. (1998) take $\psi(\boldsymbol{x})$ to be a linear combination of additional regressors $\{v_j(\boldsymbol{x})\}$, that is, functions other than the elements of $\boldsymbol{f}(\boldsymbol{x})$. Notz (1989) places a prior distribution on possible departures from the assumed model, and Rychlik (1987) assumes a finite-dimensional Ψ constrained by (20.5) and (20.8); both go on to specialize to polynomial responses with polynomial alternatives, a class of applications that will be discussed in more detail in Section 20.2.2. Yue and Hickernell (1999) assume that ψ comes from a reproducing kernel Hilbert space. The definition of such alternatives is quite technical, and it can be difficult to see just how broad the class is; the approach is however promising in that, in the examples given, it leads to designs affording a full exploration of the design space.

Some attempts have been made to construct designs tailored to estimates that are linear in the observations, but are not necessarily *lse's*. Karson et al. (1969) revisit the setup of Box and Draper (1959) and propose linear estimates that minimize the integrated squared bias (*isb*) resulting from fitting a low-order polynomial when the true response function contains higher order terms. For such estimates, they go on to propose designs that minimize the integrated variance. Draper and Sanders (1988) continue the design aspects of this approach and search for parsimonious designs meeting these criteria. Sacks and Ylvisaker (1984) study designs, robust against certain very broad classes Ψ of nonparametric alternatives, to be used with linear estimates. They comment that "...when Ψ is an infinite dimensional class ... it is by no means certain that the optimal choice of both design and linear estimates leads to least squares nor is there much a priori justification for adherence to least squares". Their approach is summarized in Chang and Notz (1996, Section 3.2) and extended in Tang (1993).

20.2.1 Minimum Bias Designs: Uniformity

In the development of Box and Draper (1959) described earlier, a motivating goal was the maximization of the power of a test of lack of fit (*lof*); this led in a natural way to considerations of bias as well as variance, thence to the prescription of choosing a *near uniform* design if consideration focusses on bias alone. Subsequent researchers have been led by this to investigate various optimality properties of the design with continuous uniform measure $\lambda(\boldsymbol{x})$, that is, Lebesgue measure on the design space $\mathcal{X}$, normed to have unit total mass. Implicit here is an assumption that the optimality will carry over to appropriate discretizations of this uniform measure; exactly how this should be done seems to be an area where further investigation could be fruitful (but see Fang and Wang 1994).

To describe the current situation, recall that the usual test of *lof* takes replicates at each of several locations, with covariates $\{\boldsymbol{x}_j\}$, and compares the unstructured *full* model $E\left[y_j\right] = \mu_j$ with the *restricted* model $\mu_j = \boldsymbol{f}'\left(\boldsymbol{x}_j\right)\theta$ being fitted. If the class of full models is written in the form (20.2), with θ defined by (20.3), then the power of the F-test of *lof* is an increasing function of the noncentrality parameter $\left(n/\sigma_\varepsilon^2\right)\mathcal{B}(\psi, \xi)$, where

$$\mathcal{B}(\psi, \xi) = \int_{\mathcal{X}} \left\{E\left[\hat{y}(\boldsymbol{x}) - y(\boldsymbol{x})\right]\right\}^2 \xi(d\boldsymbol{x}) = \int_{\mathcal{X}} \psi^2(\boldsymbol{x})\, \xi(d\boldsymbol{x}) - \boldsymbol{b}'_{\psi,\xi} \boldsymbol{M}_\xi^{-1} \boldsymbol{b}_{\psi,\xi}.$$

Maximizing the power is equivalent to maximizing $\mathcal{B}(\psi, \xi)$. This goal is closely related to the observation that the expectation of the regression-based estimate of σ_ε^2 – the mean square of the residuals – is given by $\sigma_\varepsilon^2 + n\mathcal{B}(\psi, \xi)/(n-p)$. We note as well that the *isb* of $\hat{y}(\boldsymbol{x})$ in the full model is, as at (20.7),

$$\textsc{isb}(\psi, \xi) = \int_{\mathcal{X}} \left\{E\left[\hat{y}(\boldsymbol{x}) - y(\boldsymbol{x})\right]\right\}^2 d\boldsymbol{x} = \int_{\mathcal{X}} \psi^2(\boldsymbol{x})\, d\boldsymbol{x} + \boldsymbol{b}'_{\psi,\xi} \boldsymbol{M}_\xi^{-1} \boldsymbol{A} \boldsymbol{M}_\xi^{-1} \boldsymbol{b}_{\psi,\xi}.$$

Now denote by Ψ^- the class given by (20.5) and (20.8) and by Ψ^+ the class that has the inequality reversed in (20.8). For testing *lof*, which we now express through the null hypothesis that $\psi \equiv 0$, it is necessary to separate the null and alternate hypotheses; hence, we restrict to Ψ^+. In Wiens (1991), it is shown that the uniform design has a *maximin* property – it maximizes the minimum (over Ψ^+) power of the test of *lof*, with $\mathcal{B}(\psi, \lambda) = \int_{\mathcal{X}} \psi^2(\boldsymbol{x})\, \lambda(d\boldsymbol{x}) \geq \tau_0^2 \stackrel{def}{=} \tau^2 / \int_{\mathcal{X}} d\boldsymbol{x}$. The proof consists of showing that if the design ξ is nonuniform, then one can construct $\psi \in \Psi^+$ with $\mathcal{B}(\psi, \xi) \leq \tau_0^2$. The same method of proof establishes a dual, *minimax* property of λ – it minimizes the maximum (over Ψ^-) bias of the mean square of the residuals as an estimate of σ_ε^2. Similarly, λ is also a minimax (over Ψ^-) design with respect to $\textsc{isb}(\psi, \xi)$, with $\textsc{isb}(\psi, \lambda) = \int_{\mathcal{X}} \psi^2(\boldsymbol{x})\, \lambda(d\boldsymbol{x})$.

Xie and Fang (2000) obtain similar optimality results in a nonparametric regression setting. They assume that $E[y(\boldsymbol{x})] = g(\boldsymbol{x})$ is to be estimated by a linear combination of the members of a set $\Psi_k = \{\psi_i\}_{i=1}^k$ of square-integrable and functionally independent but otherwise arbitrary functions and show that the maximum, over all such approximations, of *isb* is minimized by the continuous uniform design. They study examples in which Ψ_k consists of Fourier regression functions, or Haar wavelets.

Wiens (2000) works with the class Ψ^- and allows *weighted* least squares (*wls*) estimates (while continuing to assume homoscedastic errors), with weights $w(\boldsymbol{x})$. The analogues of $\boldsymbol{b}_{\psi,\xi}$ and $\boldsymbol{M}_\xi$ at (20.6) are $\boldsymbol{b}_{\psi,w,\xi} = \int_{\mathcal{X}} \boldsymbol{f}(\boldsymbol{x})\, \psi(\boldsymbol{x})\, w(\boldsymbol{x})\, \xi(d\boldsymbol{x})$ and $\boldsymbol{M}_{w,\xi} = \int_{\mathcal{X}} \boldsymbol{f}(\boldsymbol{x})\, \boldsymbol{f}'(\boldsymbol{x})\, w(\boldsymbol{x})\, \xi(d\boldsymbol{x})$. It follows that if ξ has a density $k(\boldsymbol{x})$ with $w(\boldsymbol{x})\, k(\boldsymbol{x}) \propto \lambda(\boldsymbol{x})$

(implying that $\boldsymbol{M}_{w,\xi} \propto \boldsymbol{A}$), then $\boldsymbol{b}_{\psi,w,\xi} \equiv 0$, that is, $E\left[\hat{\theta} - \theta \mid \psi\right] \equiv 0$. Within the class of such *unbiased* designs, a member minimizing the integrated variance of $\hat{y}(\boldsymbol{x})$ is found. This integrated variance is

$$\frac{\sigma^2}{n} \int_{\mathcal{X}} \frac{1}{w(\boldsymbol{x})} d\boldsymbol{x} \int_{\mathcal{X}} \boldsymbol{f}'(\boldsymbol{x}) \boldsymbol{A}^{-1} \boldsymbol{f}(\boldsymbol{x}) w(\boldsymbol{x}) d\boldsymbol{x},$$

minimized by weights

$$w(\boldsymbol{x}) \propto \frac{1}{\sqrt{\boldsymbol{f}'(\boldsymbol{x}) \boldsymbol{A}^{-1} \boldsymbol{f}(\boldsymbol{x})}}. \tag{20.14}$$

Examples of the resulting *minimum variance unbiased* (*mvu*) designs and weights are shown in Figure 20.4, for approximate polynomial regression.

Figure 20.4 illustrates another recurring feature that occurs not only in robust design theory but also in optimal design theory when the model is not in doubt and only the variance is minimized. This is the similarity of the *I* and *D*-optimal designs to each other and the quite different nature of the *A*-optimal designs. (In Figure 20.4 *I*-optimality is referred to as

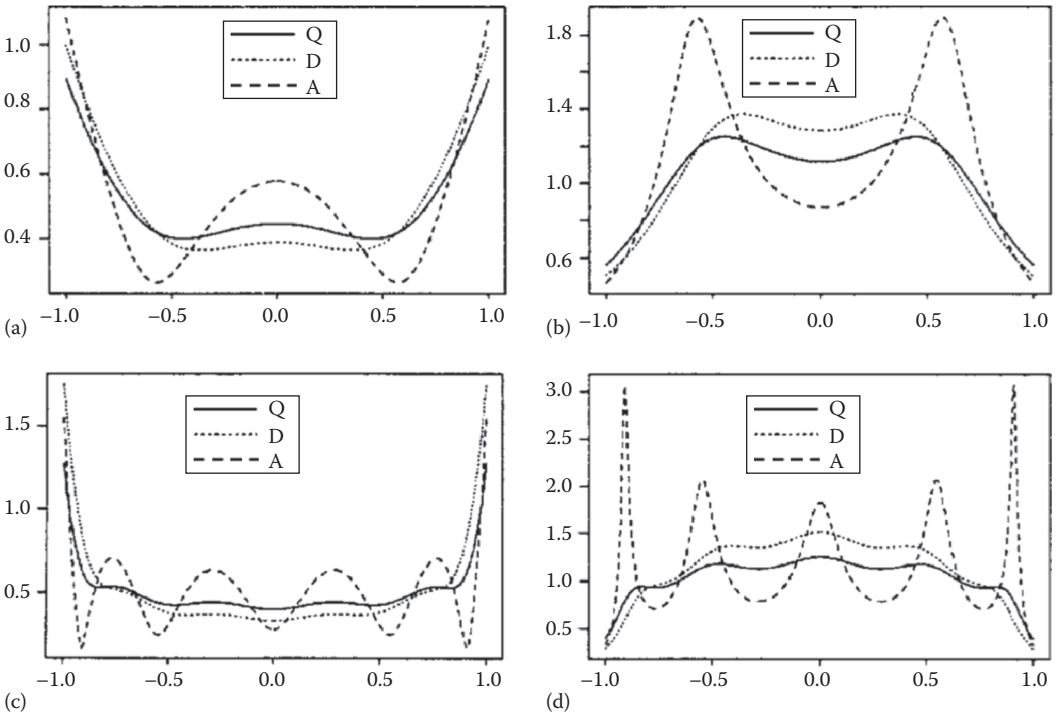

FIGURE 20.4
mvu design densities and weights for degree-*q* approximate polynomial regression and various loss functions: (a) design densities, $q = 2$; (b) weights, $q = 2$; (c) design densities, $q = 5$; (d) weights, $q = 5$. The loss functions corresponding to *Q*- (=*I*-), *D*- and *A*-optimality are respectively *imse*, the determinant of the *mse* matrix and the trace of the *mse* matrix. (From Wiens, D.P., *J. Stat. Plan. Interference*, 83, 395, 2000.)

Q-optimality, a term that is now out of fashion.) This seems to be related to the fact that the former criteria are range invariant, whereas the latter is not – a linear transformation of the design space incurs not merely a linear transformation of the *A*-optimal design points but a change in the design weights as well. See the discussion in Wierich (1988).

Biedermann and Dette (2001) extend the results of Wiens (1991, 2000) in several directions. They introduce a function $v(\boldsymbol{x})$ whose role can be characterized by replacing (20.3) by $\boldsymbol{\theta} = \arg\min_{\boldsymbol{\eta}} \int_{\mathcal{X}} \left(E[y(\boldsymbol{x})] - \boldsymbol{f}'(\boldsymbol{x})\boldsymbol{\eta}\right)^2 v(\boldsymbol{x})\, d\boldsymbol{x}$ and allow for heteroscedastic errors with variance $\sigma^2(\boldsymbol{x})$. Using the method of proof in Wiens (1991), they show that the minimum powers of each of several tests of *lof* are maximized by weights inversely proportional to $\sigma^2(\boldsymbol{x})$ and by a design density with $w(\boldsymbol{x})k(\boldsymbol{x}) \propto v(\boldsymbol{x})$. Bischoff and Miller (2006) extend Wiens (1991) and Biedermann and Dette (2001) – they propose that one design to minimize the variance, subject to a lower bound on the power of the *lof* test.

We see from this that the unique, bias-minimizing role of the uniform design is determined, through (20.3), by the experimenter's specification of just what *bias* is to mean – indeed, Biedermann and Dette (2001, p. 223) state that "...the optimality property of a particular design with respect to Wiens (1991) maximin criterion is intimately related to the particular measure used in the definition of the set Ψ". A comprehensive study of other measures used to define the target model would no doubt yield interesting unbiased designs.

20.2.2 Polynomial Models of Uncertain Degree

When the target response function being fitted is a polynomial in a single covariate x, a more focussed form of model robustness is to entertain only other polynomials, of different degrees than that being fitted. Then the matrices $\boldsymbol{M}_\xi$ under all models and the bias vector $\boldsymbol{b}_{\psi,\xi}$ (with ψ containing the powers of x of interest but not being fitted) defined in Section 20.2 depend solely on the moments $c_k(\xi) = E_\xi\left[X^k\right]$ (hence the common use of *moment matrix* to refer to $\boldsymbol{M}_\xi$ in polynomial regression). Consequently, the mathematical development is closely related to the theory of moment spaces and of *canonical moments*. Roughly speaking, the canonical moment $p_{n+1}(\xi)$ describes the possible values of $c_{n+1}(\xi)$, among all distributions ξ with given values of $c_k(\xi)$ for $1 \le k \le n$. For a book-length treatment, see Dette and Studden (1997).

Studden (1982) considered the following problem (extending a proposal of Stigler 1971). One intends to fit a polynomial of degree r to observations at $x \in [0,1]$, but wishes to be able to determine whether or not the coefficients β_{r+s} of x^{r+s} $(1 \le s \le m-r)$ would be statistically significant were these coefficients to be assessed. With $\boldsymbol{f}_1(x) = (1, x, \ldots, x^r)'$, $\boldsymbol{f}_2(x) = \left(x^{r+1}, \ldots, x^m\right)'$ and $\boldsymbol{M}_{ij}(\xi) = \int_0^1 \boldsymbol{f}_i(x)\boldsymbol{f}_j'(x)\,\xi(dx)$, the moment matrix for the model of degree m is partitioned as

$$\boldsymbol{M}_m(\xi) = \int_0^1 \begin{pmatrix} \boldsymbol{f}_1(x) \\ \boldsymbol{f}_2(x) \end{pmatrix} \left(\boldsymbol{f}_1'(x), \boldsymbol{f}_2'(x)\right) \xi(dx) = \begin{array}{c} \\ r+1 \\ m-r \end{array} \begin{array}{c} \begin{array}{cc} r+1 & m-r \end{array} \\ \begin{pmatrix} \boldsymbol{M}_{11}(\xi) & \boldsymbol{M}_{12}(\xi) \\ \boldsymbol{M}_{21}(\xi) & \boldsymbol{M}_{22}(\xi) \end{pmatrix} \end{array}. \tag{20.15}$$

Note that $\boldsymbol{M}_{11}(\xi) = \boldsymbol{M}_r(\xi)$ is the moment matrix for the fitted model of degree r. Were $\hat{\beta}_{r+1}, \ldots, \hat{\beta}_m$ to be fitted, their covariance matrix would be the inverse of $\boldsymbol{M}_{22.1} \stackrel{def}{=} \boldsymbol{M}_{22} - \boldsymbol{M}_{21}\boldsymbol{M}_{11}^{-1}\boldsymbol{M}_{12}$, and so an extended *D*-optimality criterion is to maximize $\det(\boldsymbol{M}_{11}(\xi))$, subject to a lower bound

$$\det\left(\boldsymbol{M}_{22.1}\right) \geq \rho^{m-r} \sup_{\eta} \det\left(\boldsymbol{M}_{22.1}\left(\eta\right)\right), \tag{20.16}$$

for given $\rho \leq 1$. Studden (1982) phrases and solves this problem in terms of the canonical moments and then finds the design ξ_{rm} with the optimal canonical moments. For instance, ξ_{12}, for fitting a straight line with a quadratic alternative, is of the form

$$\xi = \left\{\left(0, \alpha\right), \left(1/2, 1-2\alpha\right), \left(1, \alpha\right)\right\}, \tag{20.17}$$

with $\alpha = \left(1+\sqrt{1-\rho}\right)/4$.

Dette and Studden (1995) consider designs for polynomial regression when the experimenter seeks reasonable efficiency for all degrees $r = 1, 2, \ldots, K$. They define the *D*-efficiency

$$\mathit{eff}_r\left(\xi\right) = \left(\frac{\det\left(\boldsymbol{M}_r\left(\xi\right)\right)}{\sup_{\eta} \det\left(\boldsymbol{M}_r\left(\eta\right)\right)}\right)^{1/r+1}. \tag{20.18}$$

They then seek designs maximizing the weighted p-norm

$$\Phi_{p,\boldsymbol{w}} = \left(\sum_{r=1}^{K} w_r \left(\mathit{eff}_r\left(\xi\right)\right)^p\right)^{1/p}, \quad -\infty \leq p \leq 1, \tag{20.19}$$

for nonnegative weights $w_1, \ldots, w_K$. Again the problem is first formulated and solved in terms of canonical moments. For instance, suppose that $K = 2$ (linear or quadratic regression), $x \in [0,1]$ and the weights are uniform: $(w_1, w_2) = (1/2, 1/2) \stackrel{def}{=} \boldsymbol{w}_0$. Then the $\Phi_{p,\boldsymbol{w}_0}$-optimal design maximizing (20.19) is of the form (20.17) with α depending on p. In fact for each $p \in [-\infty, 1]$, there is $\rho \leq 1$ – recall (20.16) – such that the design of Studden (1982) is the $\Phi_{p,\boldsymbol{w}_0}$-optimal design of Dette and Studden (1995) given here. This is an instance of a more general result – that any discrete design is $\Phi_{p,\boldsymbol{w}}$-optimal for some choices of p and $\boldsymbol{w}$ – due to Dette (1991).

When the fitted polynomial response of degree m is in doubt, one might seek a high *D*-efficiency $\mathit{eff}_m\left(\xi\right)$ and simultaneously a high power for the t-test of the hypothesis that the coefficient of x^m, in a model of degree m, is zero. This latter quantity is maximized by a design maximizing the *D_1-efficiency* $\mathit{eff}_m^{D_1}\left(\xi\right)$. In the notation leading to (20.15), with $r = m-1$, the D_1-efficiencies can be defined by

$$\mathit{eff}_m^{D_1}\left(\xi\right) = \left(\frac{\det\left(\boldsymbol{M}_{22.1}\left(\xi\right)\right)}{\sup_{\eta} \det\left(\boldsymbol{M}_{22.1}\left(\eta\right)\right)}\right)^{1/m-r}.$$

(In the present case, the exponent is 1 and $\boldsymbol{M}_{22.1}$ is a scalar; we present the D_1-efficiency in this more general form for later reference.) Dette and Franke (2001) construct designs that maximize the minimum of $\left\{\mathit{eff}_m\left(\xi\right), \mathit{eff}_{m-j}^{D_1}\left(\xi\right), \ldots, \mathit{eff}_{m+k}^{D_1}\left(\xi\right)\right\}$ for nonnegative j, k. Again, canonical moments play a central role in the derivations. Fang (2006) instead maximizes a weighted average of the logarithms of $\mathit{eff}_m\left(\xi\right)$ and $\mathit{eff}_{m-1}^{D_1}\left(\xi\right)$.

Rather than considering only polynomial alternatives, one might entertain alternatives with polynomial bounds. Liu and Wiens (1997) consider degree r polynomial fits with alternatives of the form (20.2) with $\psi(x) = x^{r+1}\pi(x)$, where $\pi(\cdot)$ is continuous and bounded in absolute value by a given function $\phi(\cdot)$. Here we discuss only the case of constant $\phi(\cdot)$, so that ψ is bounded by a fixed multiple of x^{r+1}. Liu and Wiens optimize with respect to several criteria, one of which is to maximize the determinant of the moment matrix $\boldsymbol{M}_r(\xi)$, subject to a bound on the normalized bias: $\sup_\psi \boldsymbol{b}'_{\psi,\xi}\boldsymbol{M}_r^{-1}(\xi)\,\boldsymbol{b}_{\psi,\xi} \le d^2$. Designs optimal with respect to this criterion are termed *bounded-bias* (BB) designs, and it was shown that BB designs with $r+1$ support points always exist. Of course these do not allow higher order models to be explored; for instance, when $r = 1$ – straight line regression – with $x \in [0,1]$, these BB designs place equal mass at two points $\{c, 1-c\}$, with c depending on d. Fang and Wiens (2003) employ the theory of canonical moments to extend the notion to generalized bounded bias (GBB) designs. They find that if d^2 is less than the value of $\sup_\psi \boldsymbol{b}'_{\psi,\xi_D}\boldsymbol{M}_r^{-1}(\xi)\,\boldsymbol{b}_{\psi,\xi_D}$ attained by the D-optimal design ξ_D for fitting a polynomial of degree r, then the BB criterion does not determine all of the canonical moments and that more design points may then be added. They do this so as to satisfy the additional requirement of maximizing $eff_m^{D_1}(\xi)$ for a fixed $m > r$. For instance, if $r = 1$ and $m = 2$, the GBB designs on $[0,1]$ are of the form (20.17) with $\alpha = d/2 < 1/2$. If $r = 2$ and $m = 3$, there are 4 symmetrically placed design points, including 0 and 1.

20.2.3 Models with Truncated Series Response Functions

Polynomial models sometimes arise by truncating a Taylor series expansion of the response function. This interpretation is particularly appealing when the fitted terms $\boldsymbol{f}(x)$ in (20.2) constitute the first terms in a series $\sum_{i=1}^{\infty}\theta_i f_i(x)$ of orthogonal functions (over $\mathcal{X}$ and with respect to some weight function); then $\psi(x)$ may be viewed as representing the tail of the series. Oyet and Wiens (2000, 2003) and Oyet (2002) obtain minimax designs, minimizing (20.11) and its discrete or continuous analogues for loss functions such as the trace or determinant of the *mse* matrix, when the elements of $\boldsymbol{f}(x)$ are Haar wavelets or multiwavelets. For these functions and with $\mathcal{X} = [0,1]$, $\int_{\mathcal{X}} f_i(x)f_j(x)dx = \delta_{i=j}$ and then *mvu* designs and weights assume a particularly simple form: from (20.14), we obtain design densities $k(x) \propto \|\boldsymbol{f}(x)\|$ and weights $w(x) \propto 1/\|\boldsymbol{f}(x)\|$. See Figure 20.5.

Practitioners using orthogonal series expansions sometimes avoid doing least squares, since the matrix whose inversion is required can be huge. Instead, *direct estimation* is

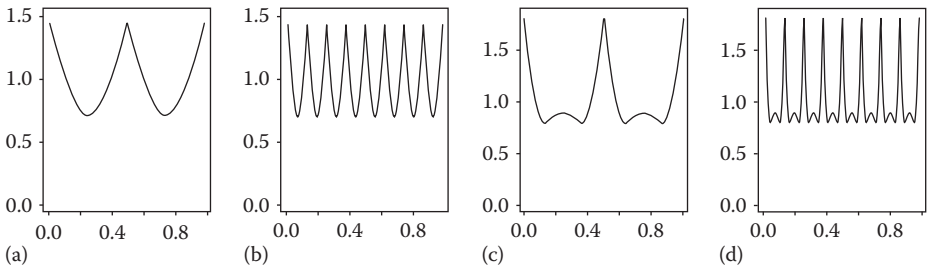

FIGURE 20.5
Minimum variance unbiased design densities for multiwavelet approximations with $N \cdot 2^{m+1}$ regressors. Each density has period $2^{-(m+1)}$; the values of N and m are (a) (2,0), (b) (2,2), (c) (3,0), and (d) (3,2). (From Oyet, A.J. and Wiens, D.P., *J. Nonparam. Stat.*, 12, 837, 2000.)

employed: if $E[y(x)] = \sum_{i=1}^{p} \theta_i f_i(\boldsymbol{x})$, with $\int_{\mathcal{X}} f_i(\boldsymbol{x}) f_j(\boldsymbol{x}) \mu(\boldsymbol{x})\, d\boldsymbol{x} = \delta_{i=j}$ for some function $\mu(\cdot)$, then $\theta_j = \int_{\mathcal{X}} E[y(\boldsymbol{x})] f_j(\boldsymbol{x}) \mu(\boldsymbol{x})\, d\boldsymbol{x}$ and estimates are computed via a suitable discretization – $\hat{\theta} = n^{-1} \sum_{i=1}^{n} y(\boldsymbol{x}_i) \boldsymbol{f}(\boldsymbol{x}_i) \mu(\boldsymbol{x}_i)$. Dette and Wiens (2008) propose the modified estimate

$$\hat{\theta} = \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{f}(\boldsymbol{x}_i) \frac{y(\boldsymbol{x}_i)\, w(\boldsymbol{x}_i)\, \mu(\boldsymbol{x}_i)}{k(\boldsymbol{x}_i)},$$

where $k(\boldsymbol{x})$ is the design density and weights $w(\boldsymbol{x})$ are introduced for flexibility in meeting robustness requirements. Constant weights, with $k(\boldsymbol{x}) \propto \|\boldsymbol{f}(\boldsymbol{x})\| \mu(\boldsymbol{x})$, yield asymptotically unbiased estimates that minimize the *imse* within the class of such unbiased estimates; otherwise, both weights and designs are chosen to minimize the maximum *imse* unconditionally. Examples are given with expansions in terms of Zernike polynomials and to spherical harmonics; in the latter case, similar results are obtained for *wls* estimation in Dette and Wiens (2009). Dette and Melas (2010) study bias-minimizing designs, with the bias arising through truncation of an expansion of the response function as a series of spline functions.

20.2.4 Nonlinear Regression Models

A problem that arises immediately in considering a design for nonlinear regression is that, by the nature of nonlinear models, the common measures of loss or efficiency depend on the unknown values of the parameters. This problem is often handled by constructing a *locally optimal* design – one that is optimal only at a particular value θ_0 of the parameter. The designer hopes that the optimality will extend to nearby parameter values. The choice of θ_0 might arise from the experimenter's prior knowledge, or perhaps as an estimate from an earlier experiment.

A mild form of robustness allows for uncertainty about the parameter values but not the form of the response function. A common approach is to first maximize the chosen loss function (or minimize an efficiency measure such as the determinant of the information matrix) over a neighbourhood of a local parameter θ_0 and to then optimize over the class of designs. In some cases, there are easily checked first-order conditions, to verify if a design is minimax. King and Wong (2000) use this approach to construct minimax *D*-optimal designs for the two-parameter logistic model and give a numerical algorithm for computing the designs; depending on the ranges of the parameters in their examples, these have up to 9 design points. Dette and Biedermann (2003) consider instead the two-parameter Michaelis–Menten model, with $E[y|x,\theta] = \theta_0 x / (\theta_1 + x)$. This model is conditionally linear in θ_0, with the consequence that the determinant of the information matrix (computed under a normal likelihood) of the *lse*'s of θ_0 and θ_1 depends only on θ_1. They find designs that maximize the minimum value of the corresponding *D*-efficiency, which is (20.18) with $r = 2$ and with $\boldsymbol{M}_2(\xi)$ replaced by the information matrix. The resulting designs are supported on two points. Dette et al. (2003) carry out a similar program for this Michaelis–Menten model, concentrating on *E*-optimality rather than *D*-optimality; in some cases, the designs are supported on three, rather than two, points. For various exponential models, see Dette et al. (2006) and Dette and Pepelyshev (2008).

Another manner in which one can deal with this problem is through Bayesian optimality. Here the loss function is averaged, with respect to a *prior* distribution on the parameters; this averaged value is then minimized. We mention Chaloner and Larntz (1989) and Dette and Neugebauer (1997) as representative approaches. We note however that at this point,

no particular viewpoint on Bayes inference is required; the adoption of a prior may be seen merely as a convenient means to an end.

The minimax and Bayesian designs described here often have very few design points, either as a consequence of the criterion or because the mathematical difficulties preclude a search for optimal designs with more support points than regression parameters. Moreover, as emphasized by Ford et al. (1989, p. 54), "Of more serious concern is the misspecification of the model itself. Some static designs offer little or no scope for testing the validity of the assumed model. Indeed, if the model is seriously in doubt, the forms of design which we have considered may be completely inappropriate." To address this issue, Sinha and Wiens (2002) entertain a class of nonlinear models forming a neighbourhood of that which the experimenter will fit, maximize the loss over this neighbourhood and then sequentially construct minimizing designs. The model development parallels (20.2) through (20.5), but with both $\boldsymbol{f}(\boldsymbol{x}|\theta) = \partial E[y|\boldsymbol{x},\theta]/\partial\theta$ and $\psi(\boldsymbol{x}|\theta)$ depending on the parameter; thus, robustness is being sought against alternate response surfaces tangent, at the parameter defined by (20.3), to that being fitted. Then (20.7) defines $\text{IMSE}(\xi|\psi,\theta)$ under homoscedasticity, with alternate forms applied if the design space is discrete or if the errors are heteroscedastic. Sinha and Wiens (2002) suppose that n observations have been made, defining the design ξ_n and yielding an estimate $\hat{\theta}_n$. They estimate ψ by smoothing the residuals, and choose the next design point, or batch of design points, so as to minimize $\text{IMSE}\left(\xi|\psi,\hat{\theta}_n\right)$, where now ξ denotes ξ_n augmented by the potential additional points. The result of a simulation study is shown in Figure 20.6. The fitted response function is exponential: $E[y|x,\theta] = \theta_0\left(1-e^{-\theta_1 x}\right)$; the true response function used in the simulations is either this exponential or the nearby Michaelis–Menten, both for $x \in [.5,5]$. Comparisons are made to the uniform and locally D-optimal design (with the local parameter being based on a small initial design). See also Sinha and Wiens (2003) for the asymptotic theory of this sequential approach.

If one insists on static designs with robustness against model misspecifications of this form, then local optimality combined with a minimax approach is a possibility; Xu (2009a) finds designs minimizing $\max_\psi \text{IMSE}(\xi|\psi,\theta_0)$ for a local parameter θ_0. One can instead take a Bayesian approach as described earlier, leading to the requirement that one minimize $\int_\Theta [\max_\psi \text{IMSE}(\xi|\psi,\theta)]\, p(\theta)\, d\theta$ – some details are in Karami and Wiens (2011), who use a discrete design space and show that $\max_\psi \text{IMSE}(\xi|\psi,\theta)$ is as at (20.13), evaluated at $\boldsymbol{f}(\cdot|\theta)$. The minimization is carried out using a genetic algorithm.

20.2.5 Generalized Linear Models

Generalized linear models (GLMs) form an increasingly popular class, for which some classical design issues were surveyed by Khuri et al. (2006). In the standard setup, it is assumed that observations are made on a random variable y, with density $p(y;\alpha,\phi) = \exp\left\{\frac{\alpha y - b(\alpha)}{a(\phi)} + c(y,\phi)\right\}$. For a vector $\boldsymbol{x}$ of covariates and vector $\boldsymbol{f}(\boldsymbol{x})$ of regressors, the mean response $E[y|\boldsymbol{x}] = \mu(\boldsymbol{x})$ satisfies $g(\mu(\boldsymbol{x})) \stackrel{def}{=} \eta(\boldsymbol{x}) = \boldsymbol{f}'(\boldsymbol{x})\,\theta$ for a strictly monotonic, differentiable function g. Some robustness issues that might be addressed at the design stage are then misspecifications of $\eta(\boldsymbol{x})$ and uncertainties about the form of $g(\cdot)$ or $a(\cdot)$. As in nonlinear regression, a further concern is the dependence of the measures of loss on the unknown parameters. Abdelbasit and Placket (1983) discuss robustness of locally optimal designs to poorly chosen local values in binary models. Again in a context of binary models, King and Wong (2000) propose maximin procedures, akin to those discussed in Section 20.2.4, with the min taken over a range of parameter values.

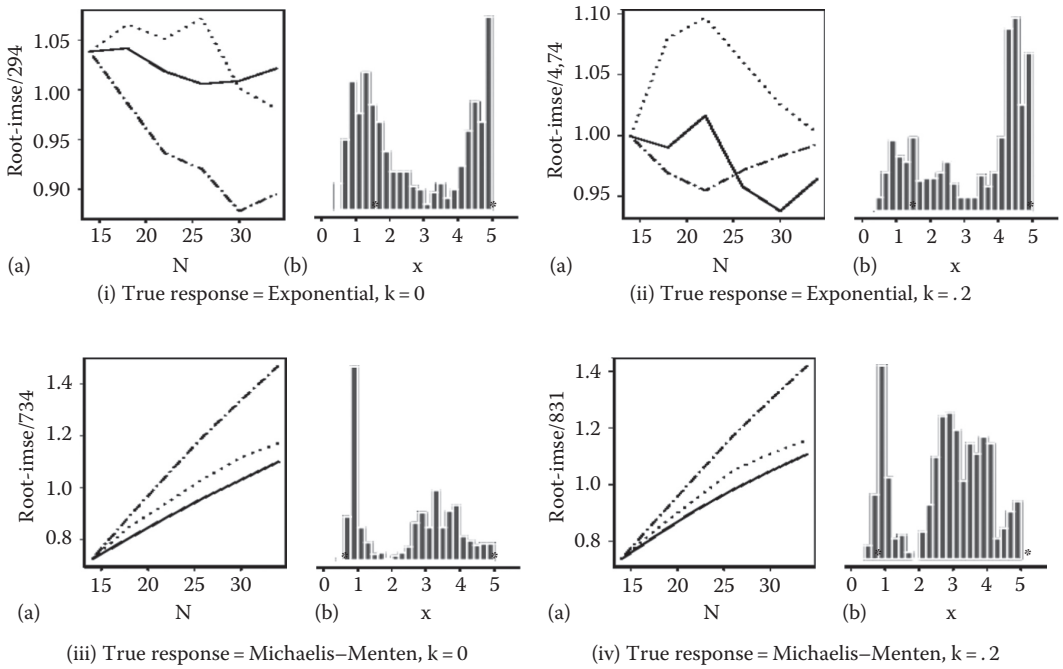

FIGURE 20.6
Sequential robust designs versus locally *D*-optimal and uniform designs. Fitted response is exponential; true response is either exponential or Michaelis–Menten. The fitting assumes homoscedasticity, whereas the true variance function is $\sigma^2(x) = 1 + k(x - .5)^2$ for $k = 0$ or $k = .2$ as indicated in the captions. (a) Average (over 100 sample paths) normalized $\sqrt{n \cdot IMSE}$ of robust (——), uniform $(\cdots)$ and *D*-optimal $(- \cdot - \cdot -)$ designs. (b) Probability histogram of all points chosen by the 100 sequential designs; asterisks are at the average sites of the *D*-optimal designs. (From Sinha, S. and Wiens, D. P., *Can. J. Stat.*, 30, 601.)

Woods et al. (2006) (see also Dror and Steinberg 2006) for a discussion of the computational aspects) seek *compromise* designs. These are defined as maximizers of the average of the log-determinant of the information matrix resulting from a design with a particular g, η and θ, with the average taken over various sets of these three quantities. Adewale and Wiens (2009) consider logistic models and robustness only against the form of $\eta(\cdot)$, which is allowed to vary over a class defined by discrete versions of (20.2) through (20.5). Adewale and Xu (2010) take an approach similar to that of Woods et al. (2006), but with the log-determinant replaced by *imse* and g replaced by a in the departures against which robustness is sought. In all of these articles, there is a mixture of analytic and numerical methods, with the latter being dominated by simulated annealing.

20.3 Robustness against a Misspecified Error Structure

20.3.1 Robustness against Heteroscedasticity

Wiens (1998) extends the methods of Section 20.2 to allow for heterogeneous error variances and allows the use of weights in the *lse's*. In the minimax procedures, a maximum is taken

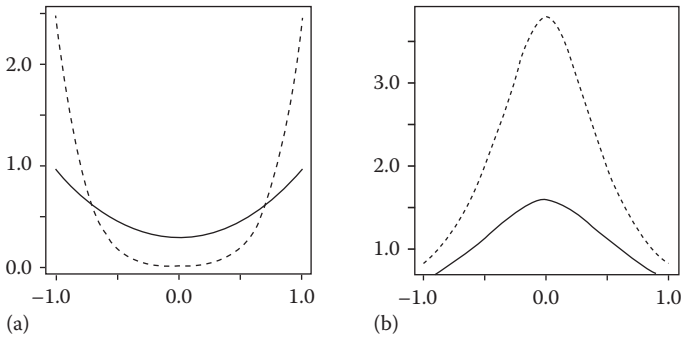

FIGURE 20.7
Minimax design densities (a) and weights (b) for heteroscedastic straight line regression for two values of ν as defined at (20.11): $\nu = 20/21$ (—) and $\nu = 1/51$ ($\cdots$). (Reprinted from Wiens, D.P., *J. Am. Stat. Assoc.*, 93, 1440, 1998. Copyright 1998 by the American Statistical Association. With permission.)

over the variance function $\sigma^2(\boldsymbol{x})$, subject to an L^2-bound, as well as the neighbourhood of regression responses, and the minimization of the resulting generalization of (20.11) yields optimal weights as well as designs. See Figure 20.7. Montepiedra and Wong (2001) suppose that the experimenter will compute the unweighted *lse*, and they investigate whether or not the *D*-optimal design continues to be optimal in the presence of a particular nonconstant variance function. They give conditions under which the two-point design for straight line regression continues to be *D*-optimal and present examples for which the *D*-optimal design that addresses the heteroscedasticity has three support points.

20.3.2 Robustness against Dependence

Experimental observations may be autocorrelated – for instance, because they are gathered serially, or because spatial or other correlations enter in through the regressors. Sacks and Ylvisaker (1966, 1968) study design problems for regression in which the error process is a time series with a known covariance function that is used in the estimation process. Bickel and Herzberg (1979) and later Bickel et al. (1981) entertain the following form of robustness. It is supposed that the data will be analyzed by ordinary least squares (*ols*), tailored to a straight line fit (with or without an intercept) in the covariate t (time). The straight line model is not in doubt, but the true autocorrelation function is a mixture of a particular autocorrelation function $\rho(\cdot)$ and that under independence, namely, $(1-\gamma)\,\delta_{\{t_i=t_j\}} + \gamma\rho\left(t_i - t_j\right)$ for a mixing parameter γ and observations made at times t_i, t_j. Their asymptotics lead to the conclusion that at least when $\rho(\cdot)$ is convex in $|t|$, the uniform design is very nearly optimal for regression with an intercept. For regression through the origin, an example of an exactly optimal design density is in Figure 20.8. From these cases, we are led to the conjecture that designs robust against autocorrelations of this form are qualitatively similar to designs constructed to be robust against misspecifications in the regression response but assuming that the observations will be independent.

Wiens and Zhou (1996) consider regression models in which the error process $\{\varepsilon(t)\}$ has a mixed autocorrelation function as the preceding, with $\rho(t)$ absolutely summable (leading to the existence of a spectral density) but otherwise arbitrary. Misspecified response functions (in covariates $\boldsymbol{x}$) following (20.2) through (20.5) are allowed and the parameters are

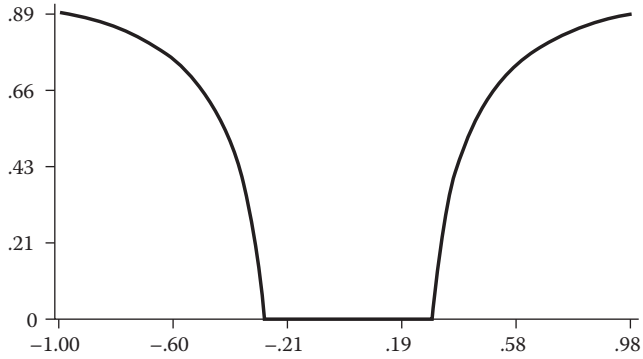

FIGURE 20.8
Optimal design density of Bickel and Herzberg (1979) for regression through the origin; $\gamma = .5$, $\rho(t) = \exp(-.483|t|)$. (From Bickel, P.J. and Herzberg, A.M., *Ann. Stat.*, 7, 77, 1979.)

estimated by *ols*. They show that *a design ξ_* that is asymptotically minimax optimal for uncorrelated errors retains its optimality if the design points come from the measure ξ_*, or from a design tending weakly to ξ_*, and are then implemented in random order*. This complements results of Wu (1981) who established the model robustness of various randomized designs for comparative experiments such as randomized complete block designs and randomized Latin square designs.

The omnibus correlation structure of the previous paragraph calls for the design points to be implemented in random order. If the class of autocorrelation models against which one seeks protection is less broad, then more precise guidance can be given. Wiens and Zhou (1997) work with models as at (20.2) through (20.5) and correlation structures as the aforementioned, expressed through autocorrelation matrices $\boldsymbol{P}_\gamma = (1-\gamma)\boldsymbol{I} + \gamma\boldsymbol{P}$ for $\boldsymbol{P} \in \mathcal{P}$. Particular attention is paid to the classes $\mathcal{P}_1$, $\mathcal{P}_2$ of MA(1) structures with positive or negative lag-1 correlations, respectively. Under $\boldsymbol{P}_\gamma$, the determinant of the *mse* matrix of the *ols* estimate $\hat{\theta}$ is

$$\mathcal{D}(\psi, \xi, \boldsymbol{P}_\gamma) = \frac{\sigma^{2p}}{n} \left|\boldsymbol{M}_\xi\right|^{-2} \left|\frac{\boldsymbol{X}'\boldsymbol{P}_\gamma\boldsymbol{X}}{n}\right| \cdot \left\{1 + \frac{n}{\sigma^2}\boldsymbol{b}'_{\psi,\xi}\left(\frac{\boldsymbol{X}'\boldsymbol{P}_\gamma\boldsymbol{X}}{n}\right)^{-1}\boldsymbol{b}_{\psi,\xi}\right\}.$$

With $\psi_0 = 0$ and $\boldsymbol{P}_0 = \boldsymbol{I}$, define the *change-of-variance function* in the direction of $\boldsymbol{P} \in \mathcal{P}$ by

$$CVF(\xi, \boldsymbol{P}) = \frac{\frac{d}{d\gamma}\mathcal{D}(\psi_0, \xi, \boldsymbol{P}_\gamma)_{|_{\gamma=0}}}{\mathcal{D}(\psi_0, \xi, \boldsymbol{P}_0)},$$

and the *change-of-bias function* in the direction of $\psi \in \Psi$ by

$$CBF(\xi, \psi) = \frac{\frac{1}{2}\frac{d^2}{d\gamma^2}\mathcal{D}((1-\gamma)\psi_0 + \gamma\psi, \xi, \boldsymbol{P}_0)_{|_{\gamma=0}}}{\sigma^{-2}\mathcal{D}(\psi_0, \xi, \boldsymbol{P}_0)}.$$

The supremum of $CVF(\xi, \boldsymbol{P})$ over $\boldsymbol{P} \in \mathcal{P}$ is termed the *change-of-variance sensitivity*, denoted $CVS(\xi, \mathcal{P})$ and that of $CBF(\xi, \psi)$ over Ψ the *change-of-bias sensitivity*, denoted $CBS(\xi, \Psi)$. A design ξ is *V-robust* if it minimizes $\mathcal{D}(\psi_0, \xi, \boldsymbol{P}_0)$, that is, maximizes $\det(\boldsymbol{M}_\xi)$, subject to

a bound $CVS\,(\xi, \mathcal{P}) \leq \alpha$, and *most V-robust* in a class of designs if α is the infimum of the *CVS* over this class. Similarly, a design is *B-robust* if it maximizes $\det\left(\boldsymbol{M}_\xi\right)$, subject to a bound $CBS\,(\xi, \psi) \leq \beta$, and *most B-robust* if β is the infimum of the *CBS* over a given class of designs. These extend notions of robustness, introduced in other contexts by Hampel et al. (1986), under which one seeks efficiency at the assumed model together with a bound on the rate at which the performance deteriorates as one moves away from this model.

It is evident that B-robust designs coincide with the BB designs of Section 20.2.2. A design is *M-robust* if it is both V-robust and B-robust and *most M-robust* if it is both most V-robust and most B-robust. The question of the existence of most M-robust designs seems to have remained open. Examples of most V-robust, most B-robust and M-robust designs are given in Wiens and Zhou (1997) for q-dimensional multiple linear regression with an intercept: $\boldsymbol{f}(\boldsymbol{x}) = \left(1, \boldsymbol{x}'\right)'$. The restriction is made to design spaces and designs for which $\boldsymbol{M}_\xi$ is diagonal. For $q = 1$ and $q = 2$, see Figure 20.9. We see from this that a general guiding principle is that the design points should be robust against response function misspecifications and then implemented in an order such that there are as many sign changes as possible under $\mathcal{P}_1$ and as few as possible under $\mathcal{P}_2$. Other examples of M-robust designs are given by Tsai and Zhou (2005).

Wiens and Zhou (1999) make similar findings, when robustness is sought against possible autocorrelations with an AR(1) structure. A method established there, to asymptotically minimize the maximum *imse*, over Ψ and over AR(1) models with lag-1 correlation $\rho < 0$, is to first generate optimal design points $\left\{\boldsymbol{t}_j\right\}_{j=1}^n$ assuming independence. Then, having implemented $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{m-1}$ (with $\boldsymbol{x}_0 = \boldsymbol{0}$), $\boldsymbol{x}_m$ is to be the nearest neighbour, among those $\boldsymbol{t}_j$ not yet chosen, of $\boldsymbol{x}_{m-1}$. If instead $\rho > 0$, then the induced design $\left\{(-1)^m \boldsymbol{x}_m\right\}_{m=1}^n$ is asymptotically minimax. Zhou (2001a) replaces the *imse* criterion by the trace of the *mse* matrix and obtains designs numerically for MA(q) errors ($q = 1, 2$), by a two-stage algorithm in which the design points are found by simulated annealing and an optimal order of implementation is obtained by a nested exchange algorithm. Shi et al. (2007) extend these results to $q = 3$.

Zhou (2001b) considers a modified form of the CVF. Motivated by a desire for accurate confidence intervals on linear functions $\boldsymbol{a}'\boldsymbol{\theta}$, she defines

$$CVF_{\boldsymbol{a}}\,(\xi, \boldsymbol{P}) = \frac{\frac{d}{d\gamma}\boldsymbol{a}'V\left(\boldsymbol{P}_\gamma\right)\boldsymbol{a}_{|_{\gamma=0}}}{\boldsymbol{a}'V\left(\boldsymbol{P}_0\right)\boldsymbol{a}},$$

where $V\left(\boldsymbol{P}_\gamma\right) = \boldsymbol{M}_\xi^{-1}\left(\frac{\boldsymbol{X}'\boldsymbol{P}_\gamma\boldsymbol{X}}{n}\right)\boldsymbol{M}_\xi^{-1}$ is the covariance matrix of the *ols* estimate $\sqrt{n}\hat{\boldsymbol{\theta}}$. It turns out, after a calculation, that

$$CVF_{\boldsymbol{a}}\,(\xi, \boldsymbol{P}) = \frac{\boldsymbol{a}'\left(V\left(\boldsymbol{P}\right) - V\left(\boldsymbol{I}\right)\right)\boldsymbol{a}}{\boldsymbol{a}'V\left(\boldsymbol{I}\right)\boldsymbol{a}},$$

the numerator of which is proportional to the difference in the squared widths of confidence intervals computed using a knowledge of $\boldsymbol{P}$ and those based on *ols*. Zhou (2001b) goes on to evaluate $CVF_{\boldsymbol{a}}\,(\xi, \boldsymbol{P})$ for fixed $\boldsymbol{P}$ and at a given design ξ, chosen, for instance, for its efficiency or robustness under independence. Then the order of implementation of the design points is optimized, so as to minimize $|CVF_{\boldsymbol{a}}\,(\xi, \boldsymbol{P})|$. An extended criterion studied is to minimize a sum $\sum_i \left|CVF_{\boldsymbol{a}_i}\,(\xi, \boldsymbol{P})\right|$ (One might instead take a maximum over $\boldsymbol{a}$, leading

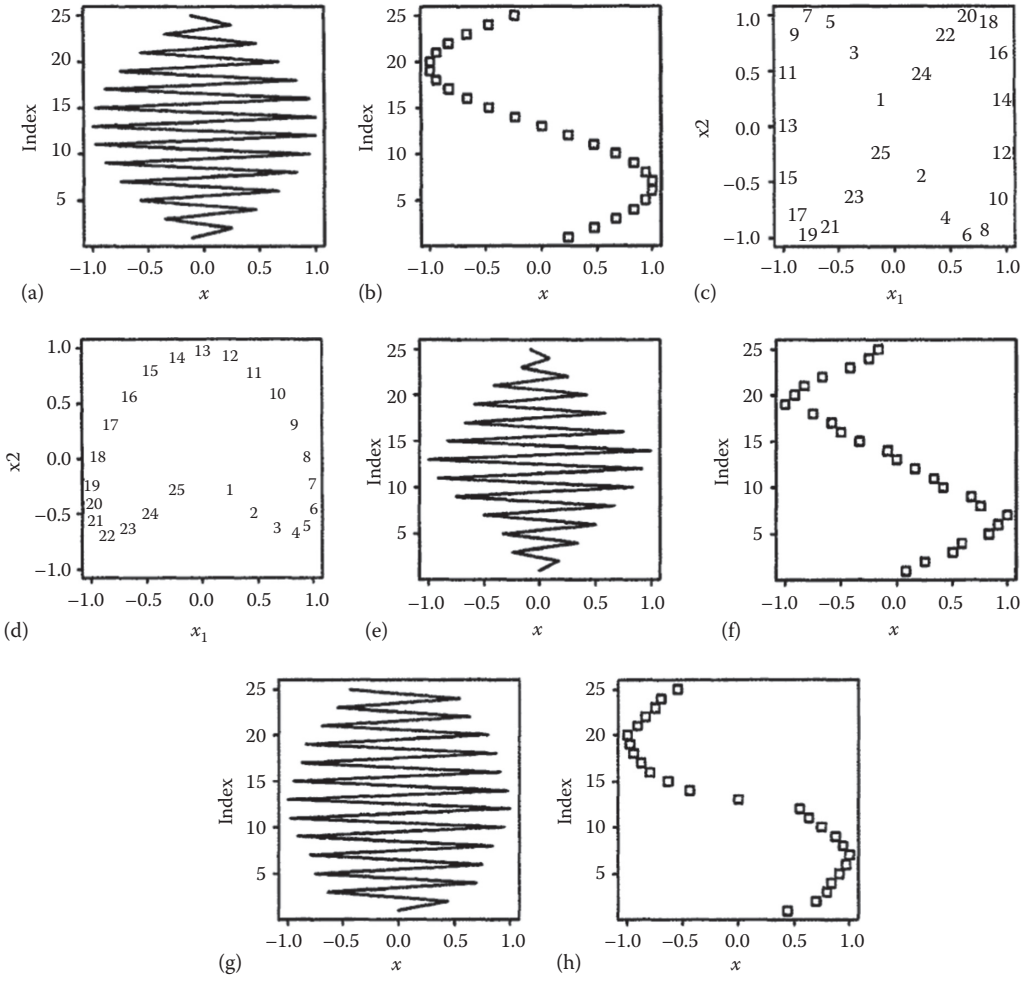

FIGURE 20.9
Most V-, most B, and M-robust designs, $n = 25$. (a,b): Most V-robust designs in $[-1, 1]$ for $\mathcal{P}_1$ and $\mathcal{P}_2$. (c,d): Most V-robust designs in $[-1, 1] \times [-1, 1]$, with the indices of the design points plotted for $\mathcal{P}_1$ and $\mathcal{P}_2$. (e,f): Most B-robust designs for Ψ, ordered for $\mathcal{P}_1$ and $\mathcal{P}_2$. (g,h): M-robust designs for Ψ, ordered for $\mathcal{P}_1$ and $\mathcal{P}_2$. In (a,e,g) the design points are at the vertices; in (b,f,g) they are at the plotted squares; in all six cases the order of implementation is as on the vertical axes. (Reprinted from Wiens, D.P. and Zhou, J., *J. Am. Stat. Assoc.*, 92, 1503, 1997. Copyright 1997 by the American Statistical Association. With permission.)

to apparently open problems.) The examples given indicate that using this optimal ordering, the coverage probabilities of confidence intervals, computed incorrectly assuming independence, deteriorate only very slightly from the nominal values.

20.4 Special Applications

In design, as in all applied research, general theory such as those presented in the previous sections must often be modified or adapted to suit specific applications. In the next few

sections, we illustrate this in the context of some particular examples in which notions of robustness of design have been found to be fruitful.

20.4.1 Extrapolation

Extrapolation poses special problems in statistics in general and design in particular, since a model that is in doubt in the design space may be even more so in the extrapolation space. Reiterating sentiments expressed by us elsewhere in this chapter, Lawless (1984) remarks that "in extrapolation problems, a slight degree of model inadequacy quickly wipes out advantages that minimum variance designs possess when the model is exactly correct".

Huber (1975) supposes that a function f is to be observed with additive random error, at n points chosen by the experimenter from $(0, \infty)$, with the intention of predicting $f(-1)$. Assuming the use of a linear predictor, he finds designs that minimize the maximum *mse* $E\left[\left\{\hat{f} - f(-1)\right\}^2\right]$, with the maximum taken over all functions f that are $h+1$ times differentiable, with a uniformly bounded $(h+1)$th derivative. These turn out to have $h+1$ points of support constituting a certain set of Chebyshev points. Huang and Studden (1988) find and correct an error in Huber's (1975) proof, while agreeing with the final result, and derive similar designs for extrapolation from $[-1, 1]$ to a single point outside of this interval. In this latter context, Hoel and Levine (1964) derive designs for extrapolating a polynomial of known degree; Spruill (1985) and Dette and Wong (1996) discuss the use of these designs, and suggest improvements, when the form or degree of the function being extrapolated is in doubt.

Fang and Wiens (2000) consider the extrapolation of approximate regression responses as at (20.2) through (20.5), but with $\boldsymbol{x}$ chosen from a finite design space $\mathcal{X}$ to a space $\mathcal{T}$ on which $E[y(\boldsymbol{x})] = \boldsymbol{f}'(\boldsymbol{x})\,\theta + \psi_T(\boldsymbol{x})$ with $\int_{\mathcal{T}} \psi_T^2(\boldsymbol{x})\,\mu(d\boldsymbol{x}) \leq \tau_T^2$ for some measure μ (typically either Lebesgue measure or counting measure). Designs minimizing the maximum, over both ψ at (20.4) and over ψ_T, of the integrated mean squared prediction error $\int_{\mathcal{T}} E\left[\left\{\hat{y}(\boldsymbol{x}) - E[y(\boldsymbol{x})]\right\}^2\right]\mu(d\boldsymbol{x})$ are derived. The development largely parallels the derivation and minimization of (20.13); in particular the maximum is derived analytically and minimized by simulated annealing. The results complement Fang and Wiens (1999) and Wiens and Xu (2008a,b), where absolutely continuous designs are sought.

To illustrate the use of these designs, consider the following experiment of Guess et al. (1977), discussed by Hoel and Jennrich (1979). A sample of $n = 235$ animals is to receive doses, at levels x, of a carcinogen. The mean response is approximated by a cubic polynomial in $x \in \mathcal{X} = [1, 500]$; the goal is extrapolation to the region $\mathcal{T} = \{0.5\}$ where the *background effect* is to be estimated. The minimum variance design proposed by Hoel and Jennrich places 63, 125, 35 and 12 observations at $x = 1$, 82.6, 342 and 500. The Huber/Huang/Studden design discussed earlier returns design weights that must be rounded to yield implementable designs; using the rounding mechanism of Pukelsheim and Rieder (1992) results in frequencies of 232, 2, 1 and 0 at $x = 1$, 125.75, 373.3 and 500 (so that the rounded design becomes supersaturated for the cubic model). With $\mathcal{X}$ discretized to 47 equally spaced points, the designs of Fang and Wiens (2000) are shown in Figure 20.10. Both homogeneous and heterogeneous error variances (ranging over a neighbourhood as described in Section 20.3.1 in the latter case), and *ols* and *wls* estimates, are considered. To some extent these designs again illustrate the paradigm of Section 20.2, with the replicates of the Hoel and Jennrich design spread out into clusters.

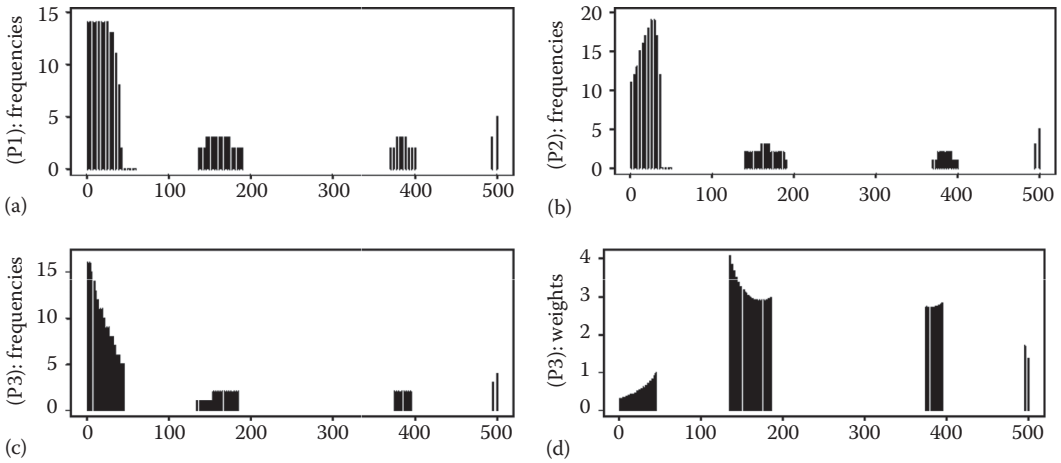

FIGURE 20.10
Minimax robust designs ($n = 235$) for extrapolation from [1, 500] to {.5}. (a) Design frequencies for *ols*, homoscedasticity. (b) design for *ols*, heteroscedasticity. (c) design for *wls*, heteroscedasticity. (d) Weights for *wls*, heteroscedasticity. (Reprinted from Fang, Z. and Wiens, D.P., *J. Am. Stat. Assoc.*, 95, 807, 2000. Copyright 2000 by the American Statistical Association. With permission.)

A class of applications in which extrapolation designs are required is accelerated life testing, in which items are tested at unusually high stress levels – at which they break down quickly, if at all – in order to predict their reliability under more usual levels. This is often combined with censoring. Chernoff (1962) derives accelerated life designs assuming that the models and parameters are known exactly, while remarking that "It is not possible to overemphasize the importance of the underlying assumptions. Doubts about them would cast serious doubts on the results to be obtained from accelerated designs". Ginebra and Sen (1998) address these reservations by constructing minimax designs, with the maximum taken over plausible ranges of the unknown model parameters. Xu (2009b) obtains designs for situations in which both acceleration and censoring are present, under departures similar to those in Fang and Wiens (2000) discussed earlier, but with a continuous design space.

20.4.2 Model Selection: Discrimination and Goodness of Fit

In the face of model uncertainty, it is natural to seek design strategies to clarify the nature of the response function. When the experimenter is choosing among a set of models whose functional forms are specified and that are perhaps even nested, then ideas similar to those in Section 20.2.2 might be applied – see, for example, Biedermann et al. (2007), where the goal is to select the correct response function from a set $\left\{\sum_{i=1}^{k} a_i \exp\left(-\lambda_i x\right); k \geq 1\right\}$ and to estimate the parameters $\{a_i, \lambda_i; i = 1, \ldots, k\}$ with reasonable efficiency. This is continued in Chapter 14.

Suppose, however, as is perhaps more common, that the correct model is a member of one of two only approximately known classes $\mathcal{C}_0, \mathcal{C}_1$. In $\mathcal{C}_j$, the response variable has density $p\left(y|\boldsymbol{x}, \mu_j\right)$, parameterized by the mean response $\mu_j\left(\boldsymbol{x}\right) = \eta_j\left(\boldsymbol{x}|\theta_j\right) + \psi\left(\boldsymbol{x}\right)$ with ψ ranging over a class Ψ_j similar to those in Section 20.2. If $\mathcal{C}_0, \mathcal{C}_1$ each contains only a single member,

that is, if each of Ψ_0 and Ψ_1 contains only the zero function, then one can discriminate between the classes by means of the Neyman–Pearson test of H_0: $p\left(y|\boldsymbol{x}, \mu_0\right)$ versus H_1: $p\left(y|\boldsymbol{x}, \mu_1\right)$. One might then seek a design maximizing the power of this test; this leads to maximizing $E_\xi\left[\mathcal{I}\left\{\mu_0\left(\boldsymbol{x}\right), \mu_1\left(\boldsymbol{x}\right)\right\}\right]$, where

$$\mathcal{I}\left\{\mu_0\left(\boldsymbol{x}\right), \mu_1\left(\boldsymbol{x}\right)\right\} = \int_{-\infty}^{\infty} p\left(y|\boldsymbol{x}, \mu_1\right) \log\left\{\frac{p\left(y|\boldsymbol{x}, \mu_1\right)}{p\left(y|\boldsymbol{x}, \mu_0\right)}\right\} dy$$

is the Kullback–Leibler (KL) divergence, measuring the information that is lost when $p\left(\cdot|\boldsymbol{x}, \mu_0\right)$ is used to approximate $p\left(\cdot|\boldsymbol{x}, \mu_1\right)$ – see López-Fidalgo et al. (2007). Wiens (2009) continues this approach and constructs *robust KL-optimal* designs that maximize the minimum value of $E_\xi\left[\mathcal{I}\left\{\mu_0\left(\boldsymbol{x}\right), \mu_1\left(\boldsymbol{x}\right)\right\}\right]$, over full classes $\mathcal{C}_0$ and $\mathcal{C}_1$. This corresponds to discriminating between the members of a least favourable pair

$$\left(\mu_0^* = \eta_0\left(\cdot|\theta_0\right) + \psi_0^*\left(\cdot\right), \mu_1^* = \eta_1\left(\cdot|\theta_1\right) + \psi_1^*\left(\cdot\right)\right),$$

at which $\mathcal{C}_0$ and $\mathcal{C}_1$ are closest. Both static and sequential designs are studied. In the former case, the designs can roughly be described as placing clusters of points in regions where the two competing mean responses are farthest apart – as one might expect. See also O'Brien (1995) who gives a method of adding locations to a *D*-optimal design – typically with no more design points than parameters – in such a way as to allow testing for *lof*.

In the sequential case, this KL criterion of Wiens is modified so as to allow a transition, over time, from choosing design points solely to maximize the discriminatory power to choosing them for increased efficiency of parameter estimation in the more plausible model. This criterion of optimizing a mixture of goals is used as well in Wiens (2010), where the KL criterion is applied to testing an approximately specified model for *lof*, while simultaneously aiming for some efficiency in estimating the parameters. This is also the goal of Dette et al. (2005), who consider the EMAX class of response functions $E\left[y|x\right] = \theta_0 x^h / \left(\theta_1 + x^h\right)$ and design to test the fit of a Michaelis–Menten response ($h = 1$).

20.4.3 Dose–Response Designs

Consider an experiment in which independent binary responses are to be observed, each representing the presence or absence of a particular response to an application of a *dose* at level x. The analyst fits a model $\Pr\left(y = 1|x\right) = F_0\left(\alpha + \beta x\right)$ for some distribution function (*link*) $F_0\left(\cdot\right)$, typically the logistic or the normal. Often the goal is the estimation of ED_{100p} – the level $x_p = \left(F_0^{-1}\left(p\right) - \alpha\right)/\beta$ at which $100p\%$ of the population will respond. The associated design problem is to determine the optimal doses and the number of observations to be made at each. This is then an instance of the class of problems discussed in Section 20.2.5, with issues related to poor initial parameter estimates and inappropriately chosen links.

Sitter (1992) notes the lack of robustness to poor initial parameter estimates and proposes minimax procedures, with the maximum taken over a given parameter region.

Only symmetric designs are considered; this restriction is dropped by Biedermann and Dette (2005). Hedayat et al. (1997, 2002) instead derive designs that are *D*-optimal among those with k ($3 \leq k \leq 7$) levels and note that as k increases, the effect of poor initial guesses for α and β decreases.

Li and Wiens (2011) design to minimize the asymptotic *mse* of the estimate $\hat{x}_p$, possibly integrated over an interval of values of p, in the face of uncertainly about the link. The estimate $\hat{x}_p$ is computed under F_0, and the *mse* is then computed under an alternate link F_n and maximized over a Kolmogorov neighbourhood $\left\{F_n | \sup |F_n(t) - F_0(t)| \leq \tau/\sqrt{n}\right\}$. Designs are computed that minimize the maximized *mse*.

To compare some of these approaches, consider an experiment of Rosenberger and Grill (1997), the aim of which was to elicit information about the relationship between stimulus level (x) and response, by estimating quantiles of the stimulus–response curve. Subjects sequentially received *marking stimuli* (auditory marking clicks) at various levels and at random times near to those of a certain event and were then asked whether the event occurred before or after the stimulus. The response y was binary, with $y = 1$ being recorded if the subject reported that the event occurred before the stimulus. The principal goal was to estimate the median (ED_{50}) of the stimulus–response curve. A secondary goal was to design so as to allow for the estimation of other quantiles such as the lower and upper quartiles. The investigators assumed a logistic link relating $\Pr(y = 1)$ to a linear function of the stimulus level. This experiment was subsequently redesigned by Biedermann, Dette and Pepelyshev (2006, henceforth BDP), by Zhu and Wong (2000, henceforth ZW) and by Li and Wiens (2011, henceforth LW). The approaches were, respectively, maximin optimality (with the minimum over a range of parameters), Bayesian optimality and minimax robustness as described in the previous paragraph. Each group constructed one design for estimation of the median and another for estimation of the quantiles.

See Figure 20.11, where the four designs of BDP and ZW are illustrated, together with four LW designs – two assuming that the parameter τ of the Kolmogorov neighbourhood defined earlier is zero, that is, that the fitted link is exactly correct, and two using $\tau = 1$. Comparative *mse* (maximized, if $\tau > 0$) in these neighbourhoods are given in Table 20.1. The designs LW2, LW4 do very well in the case of quartile estimation for which they are intended; the minimax design LW4 performs well overall. Design LW1 = LW3 also does well for ED_{50} estimation, but has a rather counterintuitive concentration of mass near the centre of the design space, illustrating that efficiency or robustness for one particular feature might lead to poor performance in other instances. An indicator of the robustness of the LW designs is that in three of the four cases in which they are intended to be optimal, there is only a small increase in *mse* when τ changes from 0 to 1.

To do only these comparisons is unfair, since we are looking at criteria with respect to which the designs LW have been optimally tuned. Those of ZW and BDP were tuned for optimal efficiency at one or several fitted models, a natural measure of which is the root of the total variance, that is, $\left[\det\left(\boldsymbol{M}_\xi\right)\right]^{-1/2}$ where $\boldsymbol{M}_\xi$ is the information matrix, given in the final column of Table 20.1. With respect to this measure, the latter designs fare well. The design LW1 = LW3 does poorly, while the others are quite comparable.

An intriguing extension of these ideas, which to our knowledge has not been pursued, is to allow $\Pr(y = 1)$ to depend not only on the level x but as well on covariates – perhaps in an only approximately specified manner. For instance, aircraft manufacturers attempt to estimate and then reduce the probability of an *ignition event* ($y = 1$), in response to a lightning strike at level x. The reduction in this probability is brought about through various surface coatings of the aircraft; these form the covariates.

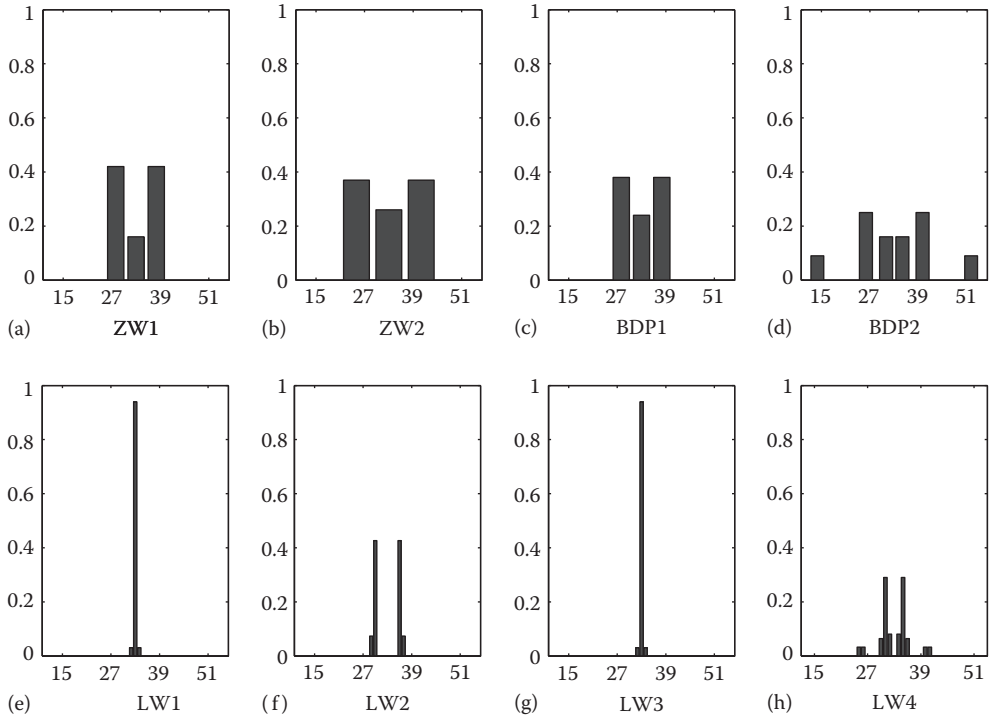

FIGURE 20.11
Dose–response designs for estimating the median (a, c, e, and g) or the quantiles (b, d, f, and h) of the stimulus–response curve. (From Li, P. and Wiens, D.P., J. R. *Stat. Soc. (Ser. B)*, 17, 215, 2011.)

TABLE 20.1

Comparative Losses of Dose–Response Designs

Design	Max mse;	ED_{50} Estimation	Max mse;	$ED_{25,75}$ Estimation	$[\det(M_\xi)]^{-1/2}$
	$\tau = 0$[a]	$\tau = 1$	$\tau = 0$[1]	$\tau = 1$	
ZW1	6.51	46.32	[3.86]	[28.33]	1.51
ZW2	[9.01]	[57.32]	5.08	49.91	1.75
BDP1	6.15	41.81	[3.74]	[26.04]	1.55
BDP2	[8.83]	[18.99]	5.12	27.85	1.91
LW1	**4.01**	4.03	[117.87]	[129.06]	16.48
LW2	[5.18]	[92.74]	**3.49**	25.06	1.65
LW3	4.01	**4.03**	[117.87]	[129.06]	16.48
LW4	[4.89]	[5.75]	3.75	**4.37**	1.93

[a] When $\tau = 0$, the loss is due solely to variation at the fitted model.

Notes: 1. Square brackets denote estimation situations in which the indicated design is not intended to be appropriate.

2. Figures in bold correspond to situations in which the LW designs are intended to be optimal.

20.4.4 Clinical Trials

In their simplest form, clinical trials involve the random allocation of patients into groups, after which each receives a treatment (perhaps a control) specific to the group. An element of randomness is required, in order that the investigator remain *blinded* to the assignments. Responses are assumed to depend on the treatment and perhaps on various covariates (age, weight, etc.). Heckman (1987) assumes that for each of two treatments ($t = 0, 1$), with a single covariate, the mean response is approximately linear:

$$E[y|x, x_0] = \beta_0(x_0) + \beta_1(x_0)\, t + \beta_2(x_0)(x - x_0) + r(t, x_0, x),$$

for $|r(t, x_0, x)|$ bounded by a known multiple, depending on t and x_0, of $(x - x_0)^2$. (This is her *model 2*; *model 1* has $\beta_2(x_0) \equiv 0$ and $|r(t, x_0, x)|$ is bounded by a multiple of $|x - x_0|$.) The difference in the mean effects of the treatments at $x = x_0$ is $\beta_1(x_0)$, so that one is interested in estimating this quantity. Sacks and Ylvisaker (1978) had obtained linear, minimum *mse* estimators for such models, and Heckman obtains the correspondingly optimal sequences of allocations to the two groups. Wiens (2005a) instead takes

$$E[y|\boldsymbol{x}] = \beta_0 + \beta_1 t + \boldsymbol{f}'(\boldsymbol{x})\,\theta + \psi_t(\boldsymbol{x})$$

(his *model 2*; *model 1* has $\theta \equiv \mathbf{0}$). Static and sequential allocation schemes are considered; in the former case a function $\rho(\cdot)$ is derived, and a subject who arrives with covariates $\boldsymbol{x}$ is assigned to the treatment group with probability $\rho(\boldsymbol{x})$. For this scheme, the results are qualitatively similar to those of Heckman. In the sequential case, they compare favourably with those of Atkinson (1982) for exactly linear responses when such do indeed hold, and have significantly reduced *mse* otherwise. There is however a significant cost to be paid for this robustness in terms of *balance* (see the discussions in Atkinson 1996, 2002) across the covariates.

20.4.5 Spatial Designs: Field and Computer Experiments

In spatial experiments – covered more broadly in Chapter 15 – one makes observations on units that are spatially arranged, and the spatial layout becomes important. A fairly general formulation is that the experimenter samples locations chosen from a set $\mathcal{X} = \{\boldsymbol{t}_1, \ldots, \boldsymbol{t}_N\} \subset \mathbb{R}^d$. At these locations, he observes

$$y(\boldsymbol{t}) = \mu(\boldsymbol{t}; \theta) + \delta(\boldsymbol{t}) + \varepsilon(\boldsymbol{t}), \tag{20.20}$$

where $\mu(\boldsymbol{t}; \theta)$ is a deterministic mean function possibly depending on unknown parameters θ, $\delta(\boldsymbol{t})$ represents stochastic, spatially correlated departures, and $\varepsilon(\boldsymbol{t})$ represents measurement error. The goals often focus on the prediction of functions of $X(\boldsymbol{t}) = \mu(\boldsymbol{t}; \theta) + \delta(\boldsymbol{t})$. Such models and prediction techniques are commonly also used in computer experiments, described in Section V; designs constructed in this context often (see in particular Chapter 17) aim for a fairly uniform exploration of the design space – itself a quite robust approach, as we have seen.

Misspecification in (20.20) can occur in a number of ways – (1) misspecifying $\mu(\boldsymbol{t}; \theta)$, perhaps in ways as outlined in Section 20.2; (2) misspecifying the covariance structure of $\delta(\boldsymbol{t})$,

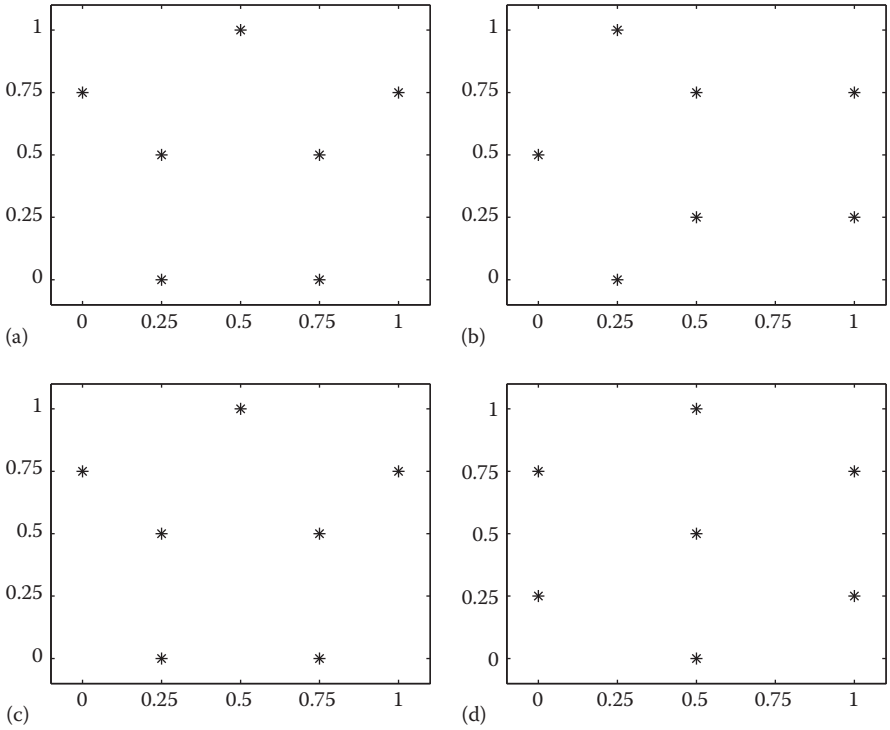

FIGURE 20.12
Minimax spatial designs for prediction of X_{Total} as in Section 20.4.5. (a) Isotropic correlations; a maximum is taken over a neighbourhood of $\mu(\boldsymbol{t};\theta)$ but the nominal correlation function for δ and homogeneous variance function for ε are assumed to be correct. (b) As in (a) but with anisotropic correlations. (c,d): As in (a,b), respectively, but maxima are taken as well over neighbourhoods of the correlation function for δ and over the nominal variance function of ε. (From Wiens, D. P., *Environmetrics*, 16, 205, 2005b.)

for instance, by assuming an isotropic structure when such is not the case; and (3) misspecifying the variance/covariance structure of $\varepsilon(\boldsymbol{t})$, as in Section 20.3. Wiens (2005b) constructs designs that are minimax robust against any of these three types of departures, using neighbourhoods of the nominal model to describe each, and simulated annealing to carry out the minimization. See Figure 20.12 for examples – in each case, 7 points are sampled from a 25-point grid (t_1, t_2). The nominal response is $\mu(\boldsymbol{t};\theta) = \theta_0 + \theta_1 t_1 + \theta_2 t_2$, the nominal covariance function Gaussian: $\text{CORR}\left[\delta(\boldsymbol{t}), \delta\left(\boldsymbol{t}'\right)\right] = \exp\left\{-\lambda_1\left(t_1 - t_1'\right)^2 - \lambda_2\left(t_2 - t_2'\right)^2\right\}$, isotropic if $\lambda_1 = \lambda_2$ and anisotropic otherwise, and the nominal error variances homogeneous. The aim is to predict $X_{Total} = \sum_{\boldsymbol{t}\in\mathcal{X}} X(\boldsymbol{t})$ with minimum *mse*.

Spatial designs are commonly employed in field experiments, where plots, perhaps in a rectangular grid, are to be planted with various crops. The goal is the optimal – with respect to some criterion – assignment of crops to plots. Minimax designs in such situations are studied by Wiens and Zhou (2008) and Ou and Zhou (2009). An example given by Ou and Zhou has eight crops, with crop one being a *control*, arranged in a 3×8 field of plots. Possible correlation structures are $NN(\rho)$, in which each plot has nonzero correlation ρ only with its immediate neighbours, and $DG(\rho)$, in which the correlation between plots is ρ^d when d is the sum of the east-west and north-south distances between them. These are the nominal

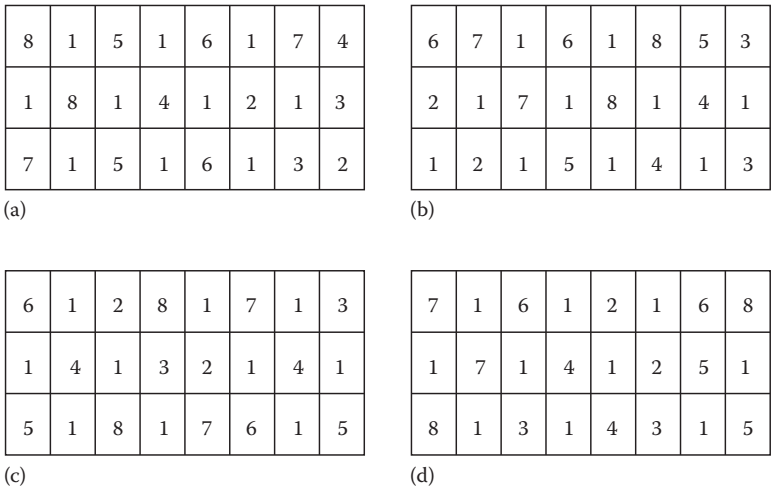

FIGURE 20.13
Minimax assignment of 8 crops to 24 fields. (a) $NN(\rho)$ correlations with *ols*, (b) $NN(\rho)$ correlations with *gls*, (c) $DG(\rho)$ correlations with *ols*, (d) $DG(\rho)$ correlations with *gls*. In all cases $\rho = .15$. (From Ou, B. and Zhou, J., *Metrika*, 69, 45, 2009.)

correlations; neighbourhood structures of these are introduced, and minimax designs are obtained assuming the use of either *ols* or generalized least squares (*gls*) using the possibly incorrect nominal correlation structure. In each case, various arrangements of replicates are compared, and it is found that, quite generally, 10 replicates of the control and 2 of all other crops is optimal for the goal of efficient estimation of the crop effects (row/column effects were not modelled). See Figure 20.13.

This brief description barely scratches the surface of applications in which spatial designers might benefit from principles of robustness. As examples, two that appear to be as yet unexplored are spatial models with discrete – perhaps binary – data, and threshold models, describing, for instance, the possible exceedance of a regulatory threshold.

20.4.6 Mixture Experiments

In mixture experiments, the covariates $x_1, \ldots, x_q$ are nonnegative and sum to one, representing the relative proportions of ingredients blended in a mixture. This is another area that is relatively unexplored by robust designers. Huang et al. (2009) find designs in the case that the experimenter is uncertain whether a first- or second-order model is appropriate. Smucker et al. (2011) extend this idea to designing so as to maximize a geometrically weighted product $\phi(\xi) = \prod_j \left[\det\left(\boldsymbol{M}_\xi\left(\boldsymbol{f}_j\right)\right)\right]^{\omega_j}$ of determinants of information matrices arising from candidate models with regressors $\boldsymbol{f}_j$. Exchange algorithms are derived for the maximization.

20.4.7 Designs for Robust Inference

Even the most carefully designed experiment can lose effectiveness if the responses themselves are erroneous, or outlying, or missing altogether. While these eventualities are not directly under the control of the designer, their effects can be tempered by an appropriate design.

The influence of an observation, at $\boldsymbol{x}_i$, on the *lse* in a linear model is measured by its *leverage* – the diagonal element h_{ii} of the *hat* matrix $\boldsymbol{X}\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'$. Recall that the h_{ii} lie in $[0, 1]$, and that their average is $\bar{h} = p/n$. Huber (1975) recommends reducing, by (approximate) replication, the values of any h_{ii} that are close to 1. Box and Draper (1975) argue that one should aim for a small value of $\sum_{i=1}^{n} h_{ii}^2$, or equivalently of the variance of the h_{ii}. They point out that this is also related to robustness against nonnormality, an observation bolstered by Huber's (1973, p. 804) remark that the condition that $\max h_{ii} \to 0$ as $n \to \infty$ "appears to be indispensable for any reasonably simple general asymptotic theory of robust estimation".

Herzberg and Andrews (1976) and Andrews and Herzberg (1979) discuss the possibility of design *breakdown*, by which they mean that the model matrix might become rank deficient after perceived outliers have been discarded, or as a result of cases that are missing entirely. This line of investigation is continued by Akhtar and Prescott (1986); Herzberg et al. (1987); and Ahmad and Gilmour (2010). They find that the loss in information (relative to the *D*-criterion) due to a missing observation at $\boldsymbol{x}_i$ is also measured by h_{ii}, and so a design with moderate values of $\{h_{ii}\}$ is robust to missing observations.

In the literature of robust estimation, one is commonly cautioned against removing cases thought to be outliers; instead, one is encouraged to downweight them in the estimation process. This is the effect of ordinary and generalized M-estimation, each of which includes least squares as a special case. Wiens and Wu (2010) construct designs that minimize the maximum asymptotic *mse* of M-estimates of regression, with the maximum taken over the discrete versions of neighbourhoods as at (20.2) through (20.5). This *mse* depends on the error distribution and on the score functions used in the estimation, but it is found that the designs change only slowly with variations in these quantities, so that a design which is robust for *lse*, against the types of departures of concern in this chapter, will remain so, if perhaps slightly suboptimal, for M-estimation.

References

Abdelbasit, K. M. and Plackett, R. L. (1983), Experimental design for binary data, *Journal of the American Statistical Association*, 78, 90–98.

Adewale, A. and Wiens, D. P. (2006), New criteria for robust integer-valued designs in linear models, *Computational Statistics and Data Analysis*, 51, 723–736.

Adewale, A. and Wiens, D. P. (2009), Robust designs for misspecified logistic models, *Journal of Statistical Planning and Inference*, 139, 3–15.

Adewale, A. J. and Xu, X. (2010), Robust designs for generalized linear models with possible overdispersion and misspecified link functions, *Computational Statistics and Data Analysis*, 54, 875–890.

Andrews, D. and Herzberg, A. (1979), The robustness and optimality of response surface designs, *Journal of Statistical Planning and Inference*, 3, 249–257.

Ahmad, T. and Gilmour, S. G. (2010), Robustness of subset response surface designs to missing observations, *Journal of Statistical Planning and Inference*, 140, 92–103.

Akhtar, M. and Prescott, P. (1986), Response surface designs robust to missing observations, *Communications in Statistics: Simulation and Computation*, 15, 345–363.

Atkinson, A. C. (1982), Optimum biased coin designs for sequential clinical trials with prognostic factors, *Biometrika*, 69, 61–67.

Atkinson, A. C. (1996), The usefulness of optimum experimental designs, *Journal of the Royal Statistical Society (Series B)*, 58, 59–76.

Atkinson, A. C. (2002), The comparison of designs for sequential clinical trials with covariate information, *Journal of the Royal Statistical Society (Series A*), 165, 349–373.

Bickel, P. J. and Herzberg, A. M. (1979), Robustness of design against autocorrelation in time I: Asymptotic theory, optimality for location and linear regression, *The Annals of Statistics*, 7, 77–95.

Bickel, P. J., Herzberg, A. M., and Schilling, M. F. (1981), Robustness of design against autocorrelation in time II: Optimality, theoretical and numerical results for the first-order autoregressive process, *Journal of the American Statistical Association*, 76, 870–877.

Biedermann, S. and Dette, H. (2001), Optimal designs for testing the functional form of a regression via nonparametric estimation techniques, *Statistics and Probability Letters*, 52, 215–224.

Biedermann, S. and Dette, H. (2005), Numerical construction of parameter maximin *D*-optimal designs for binary response models, *South African Statistical Journal*, 39, 127–161.

Biedermann, S., Dette, H., and Pepelyshev, A. (2006), Some robust design strategies for percentile estimation in binary response models, *The Canadian Journal of Statistics*, 4, 603–622.

Biedermann, S., Dette, H., and Pepelyshev, A. (2007), Optimal discrimination designs for exponential regression models, *Journal of Statistical Planning and Inference*, 137, 2579–2592.

Bischoff, W. and Miller, F. (2006), Optimal designs which are efficient for lack of fit tests, *Annals of Statistics*, 34, 2015–2025.

Box, G. E. P. (1953), Non-normality and tests on variances, *Biometrika*, 40, 318–335.

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.

Box, G. E. P. and Draper, N. R. (1975), Robust designs, *Biometrika*, 62, 347–352.

Chaloner, K. and Larntz, K. (1989), Optimal Bayesian experimental design applied to logistic regression experiments, *Journal of Statistical Planning and Inference*, 21, 191–208.

Chang, Y.-J. and Notz, W. I. (1996), Model robust designs, in: *Design and Analysis of Experiments*, *Handbook of Statistics* Vol. 13, pp. 1055–1098; eds. Ghosh, S. and Rao, C. R., North-Holland, the Netherlands: Elsevier.

Chernoff, H. (1962), Optimal accelerated life designs for estimation, *Technometrics*, 4, 381–408.

Daemi, M. and Wiens, D. P. (2012), Techniques for the construction of robust regression designs, *The Canadian Journal of Statistics*, 41, 679–695.

Dette, H. (1991), A note on robust designs for polynomial regression, *Journal of Statistical Planning and Inference*, 28, 223–232.

Dette, H. and Biedermann, S. (2003), Robust and efficient designs for the Michaelis–Menten model, *Journal of the American Statistical Association*, 98, 679–686.

Dette, H. and Franke, T. (2001), Robust designs for polynomial regression by maximizing a minimum of D- and D1-efficiencies, *The Annals of Statistics*, 29, 1024–1049.

Dette, H., Lopez, I. M., Ortiz Rodriguez, I. M., and Pepelyshev, A. (2006), Maximin efficient design of experiment for exponential regression models, *Journal of Statistical Planning and Inference*, 136, 4397–4418.

Dette, H. and Melas, V. B. (2010), A note on all-bias designs With applications in spline regression models, *Journal of Statistical Planning and Inference*, 140, 2037–2045.

Dette, H., Melas, V. B., and Pepelyshev, A. (2003), Standardized maximin *E*-optimal designs for the Michaelis-Menten model, *Statistica Sinica*, 13, 1147–1163.

Dette, H., Melas, V. B., and Wong, W.-K. (2005), Optimal design for goodness-of-fit of the Michaelis–Menten enzyme kinetic function, *Journal of the American Statistical Association*, 100, 1370–1381.

Dette, H. and Neugebauer, H.-M. (1997), Bayesian D-optimal designs for exponential regression models, *Journal of Statistical Planning and Inference*, 60, 331–349.

Dette, H. and Pepelyshev, A. (2008), Efficient experimental designs for sigmoidal growth models, *Journal of Statistical Planning and Inference*, 138, 2–17.

Dette, H. and Studden, W. J. (1995), Optimal designs for polynomial regression when the degree is not known, *Statistica Sinica*, 5, 459–473.

Dette, H. and Studden, W. J. (1997). *The Theory of Canonical Moments with Applications in Statistics, Probability, and Analysis*, New York: Wiley.

Dette, H. and Wiens, D. P. (2008), Robust designs for series estimation, *Computational Statistics and Data Analysis*, 52, 4305–4324.

Dette, H. and Wiens, D. P. (2009), Robust designs for 3D shape analysis with spherical harmonic descriptors, *Statistica Sinica*, 19, 83–102.

Dette, H. and Wong, W.-K. (1996), Robust optimal extrapolation designs, *Biometrika*, 83, 667–680.

Draper, N. R. and Sanders, E. R. (1988), Designs for minimum bias estimation, *Technometrics*, 30, 319–325.

Dror, H. A. and Steinberg, D. M. (2006), Robust experimental design for multivariate generalized linear models, *Technometrics*, 48, 520–529.

Fang, K. T. and Wang, Y. (1994), *Number-Theoretic Methods in Statistics*, New York: Chapman and Hall.

Fang, Z. (2006), Some robust designs for polynomial regression models, *The Canadian Journal of Statistics*, 34, 623–638.

Fang, Z. and Wiens, D. P. (1999), Robust extrapolation designs and weights for biased regression models with heteroscedastic errors, *The Canadian Journal of Statistics*, 27, 751–770.

Fang, Z. and Wiens, D. P. (2000), Integer-valued, minimax robust designs for estimation and extrapolation in heteroscedastic, approximately linear models, *Journal of the American Statistical Association*, 95, 807–818.

Fang, Z. and Wiens, D. P. (2003), Robust regression designs for approximate polynomial models, *Journal of Statistical Planning and Inference*, 117, 305 - 321.

Fedorov, V. V., Montepiedra, G., and Nachtsheim, C. J. (1998), Optimal design and the model validity range, *Journal of Statistical Planning and Inference*, 72, 215–227.

Ford, I., Titterington, D. M., and Kitsos, C. P. (1989), Recent advances in nonlinear experimental design, *Technometrics*, 31, 49–60.

Ginebra, J. and Sen, A. (1998), Minimax approach to accelerated life tests, *IEEE Transactions on Reliability*, 47, 261–267.

Guess, H., Crump, K., and Peto, R. (1977), Uncertainty estimates for low-dose extrapolation of animal carcinogenicity Data, *Cancer Research*, 37, 3475–3483.

Hampel, F. R., Ronchetti, E., Rousseeuw, R. J., and Stahel, W. (1986), *Robust Statistics: The Approach Based on Influence Functions*, New York: Wiley.

Heckman, N. E. (1987), Robust design in a two treatment comparison in the presence of a covariate, *Journal of Statistical Planning and Inference*, 16, 75–81.

Hedayat, A. S., Yan, B., and Pezzuto, J. M. (1997), Modeling and identifying optimum designs for fitting dose-response curves based on raw optical density data, *Journal of the American Statistical Association*, 92, 1132–1140.

Hedayat, A. S., Yan, B., and Pezzuto, J. M. (2002), Optimum designs for estimating ED_p based on raw optical density data, *Journal of Statistical Planning and Inference*, 104, 161–174.

Heo, G., Schmuland, B., and Wiens, D. P. (2001), Restricted minimax robust designs for misspecified regression models, *The Canadian Journal of Statistics*, 29, 117–128.

Herzberg, A. M. and Andrews, D. F. (1976), Some considerations in the optimal design of experiments in non-optimal situations, *Journal of the Royal Statistical Society, Series B*, 38, 284–289 (Corr: V39 p394).

Herzberg, A. M., Prescott, P., and Akhtar, M. (1987). Equi-information robust designs: Which designs are possible? *The Canadian Journal of Statistics*, 15, 71–76.

Hoel, P. G. and Jennrich, R. I. (1979), Optimal designs for dose response experiments in cancer research, *Biometrika*, 66, 307–316.

Hoel, P. G. and Levine, A. (1964), Optimal spacing and weighting in polynomial prediction, *Annals of Mathematical Statistics*, 35, 1533–1560.

Huang, M.-N. L., Hsu, H.-L., Chou, C.-J., and Klein, T. (2009), Model-Robust *D*- and *A*-optimal designs for mixture experiments, *Statistica Sinica*, 19, 1055–1075.

Huang, M.-N. L. and Studden, W. J. (1988), Model robust extrapolation designs, *Journal of Statistical Planning and Inference*, 18, 1–24.

Huber, P. J. (1964), Robust estimation of a location parameter, *The Annals of Mathematical Statistics*, 35, 73–101.

Huber, P. J. (1973), Robust regression: Asymptotics, conjectures and Monte-Carlo, *The Annals of Statistics,* 1, 799–821.

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

Huber, P. J. (1981), *Robust Statistics*, Wiley.

Karami, H. K. and Wiens, D. P. (2014), Robust static designs for approximately specified nonlinear regression Models, *Journal of Statistical Planning and Inference*, 144, 55–62.

Karson, M. J., Manson, A. R., and Hader, R. J. (1969), Minimum bias estimation and experimental design for response surfaces, *Technometrics*, 11, 461–475.

Khuri, A. I., Mukherjee, B., Sinha, B. K., and Ghosh, M. (2006), Design issues for generalized linear models: A review, *Statistical Science*, 21, 376–399.

Kiefer, J. (1975), Optimal design: Variation in structure and performance under change of criterion, *Biometrika*, 62, 277–288.

King, J. and Wong, W.-K. (2000), Minimax *D*-optimal designs for the logistic model, *Biometrics*, 56, 1263–1267.

Lawless, J. F. (1984), Some problems concerning experimental designs for extrapolation, in: *Topics in Applied Statistics*, ed. Dwivedi, T. D., Montreal, Quebec, Canada: Concordia University Press, pp. 357–366.

Li, K. C. (1984), Robust regression designs when the design space consists of finitely many points, *The Annals of Statistics*, 12, 269–282.

Li, K. C. and Notz, W. (1982), Robust designs for nearly linear regression, *Journal of Statistical Planning and Inference*, 6, 135–151.

Li, P. and Wiens, D. P. (2011), Robustness of design for dose-response studies, *Journal of the Royal Statistical Society (Series B)*, 17, 215–238.

Liu, S. X. and Wiens, D. P. (1997), Robust designs for approximately polynomial regression, *Journal of Statistical Planning and Inference*, 64, 369–381.

López-Fidalgo, J., Tommasi, C., and Trandafir, P. C. (2007), An optimal experimental design criterion for discriminating between non-normal models, *Journal of the Royal Statistical Society B*, 69, 231–242.

Marcus, M. B. and Sacks, J. (1976), Robust designs for regression problems, in: *Statistical Theory and Related Topics II*, eds. Gupta, S. S. and Moore, D. S., Academic Press, New York, pp. 245–268.

Montepiedra, G. and Wong, W.-K. (2001), A new design criterion when heteroscedasticity is ignored, *Annals of the Institute of Statistical Mathematics*, 53, 418–426.

Notz, W. (1989), Optimal designs for regression models with possible bias, *Journal of Statistical Planning and Inference*, 22, 43–54.

O'Brien, T. E. (1995), Optimal design and lack of fit in nonlinear regression models, in: *Proceedings of the 10th International Workshop on Statistical Modelling*, Lecture Notes in Statistics, Springer-Verlag, New York, pp. 201–206.

Ou, B. and Zhou, J. (2009), Minimax robust designs for field experiments, *Metrika*, 69, 45–54.

Oyet, A. J. (2002), Minimax *A*- and *D*-optimal integer-valued wavelet designs for estimation, *The Canadian Journal of Statistics* 30, 301–316.

Oyet, A. J. and Wiens, D. P. (2000), Robust designs for wavelet approximations of regression models, *Journal of Nonparametric Statistics,* 12, 837–859.

Oyet, A. J. and Wiens, D. P. (2003), On exact minimax wavelet designs obtained by simulated annealing, *Statistics and Probability Letters*, 61, 111–121.

Pesotchinsky, L. (1978), Φ_p-optimal second order designs for symmetric regions,*Journal of Statistical Planning and Inference*, 2, 173–188.

Pesotchinsky, L. (1982), Optimal robust designs: linear regression in R^k, *The Annals of Statistics*, 10, 511–525.

Pukelsheim, F. and Rieder, S. (1992), Efficient rounding of approximate designs, *Biometrika*, 79, 763–770.

Rosenberger, W. F. and Grill, S. E. (1997), A sequential design for psychophysical experiments: An application to estimating timing of sensory events, *Statistics in Medicine*, 16, 2245–2260.

Rychlik, T. (1987), Robust experimental design: A comment on Huber's result, *Zastosowania Matematyki*, 19, 93–107.

Sacks, J. and Ylvisaker, D. (1966), Designs for regression problems with correlated errors, *The Annals of Statistics*, 37, 66–89.

Sacks, J. and Ylvisaker, D. (1968), Designs for regression problems with correlated errors; Many Parameters, *The Annals of Statistics*, 39, 49–69.

Sacks, J. and Ylvisaker, D. (1978), Linear estimation in approximately linear models, *The Annals of Statistics*, 6, 1122–1137.

Sacks, J. and Ylvisaker, D. (1984), Some model robust designs in regression, *The Annals of Statistics*, 12, 1324–1348.

Shi, P., Ye, J., and Zhou, J. (2003), Minimax robust designs for misspecified regression models, *The Canadian Journal of Statistics*, 31, 397–414.

Shi, P., Ye, J., and Zhou, J. (2007), Discrete minimax designs for regression models with autocorrelated MA errors, *Journal of Statistical Planning and Inference*, 137, 2721–2731.

Sinha, S. and Wiens, D. P. (2002), Robust sequential designs for nonlinear regression, *The Canadian Journal of Statistics*, 30, 601–618.

Sinha, S. and Wiens, D. P. (2003), Asymptotics for robust sequential designs in misspecified regression models; mathematical statistics and applications, *Festschrift for Constance van Eeden, IMS Lecture Notes - in: Monograph Series*, eds. Moore, M., Léger, C. and Froda, S., 233–248. Institute of Mathematical Statistics, Hayward, CA.

Sitter, R. R. (1992), Robust designs for binary data, *Biometrics*, 48, 1145–1155.

Smucker, B. J., del Castillo, E. and Rosenberger, J. (2011), Exchange algorithms for constructing model-robust experimental designs, *Journal of Quality Technology,* 43, 1–15.

Spruill, C. (1985), Model robustness of hoel-levine optimal designs, *Journal of Statistical Planning and Inference*, 11, 217–225.

Stigler, S. M. (1971), Optimal experimental design for polynomial regression, *Journal of the American Statistical Association*, 66, 311–318.

Studden, W. J. (1982), Some robust-type *D*-optimal designs in polynomial regression, *Journal of the American Statistical Association*, 77, 916–921.

Tang, D.-I. (1993), Minimax regression designs under uniform departure models, *The Annals of Statistics*, 21, 434–446.

Tsai, Y.-L. and Zhou, J. (2005), Discrete M-robust designs for regression models, *Journal of Statistical Planning and Inference*, 131, 393–406.

Wiens, D. P. (1990), Robust, minimax designs for multiple linear regression, *Linear Algebra and Its Applications,* 127, 327 - 340.

Wiens, D. P. (1991), Designs for approximately linear regression: Two optimality properties of uniform designs, *Statistics and Probability Letters*; 12, 217–221.

Wiens, D. P. (1992), Minimax designs for approximately linear regression, *Journal of Statistical Planning and Inference,* 31, 353–371.

Wiens, D. P. (1993), Designs for approximately linear regression: Maximizing the minimum coverage probability of confidence ellipsoids, *The Canadian Journal of Statistics*, 21, 59–70.

Wiens, D. P. (1998), Minimax robust designs and weights for approximately specified regression models with heteroscedastic errors, *Journal of the American Statistical Association*, 93, 1440–1450.

Wiens, D. P. (2000), Robust weights and designs for biased regression models: Least squares and generalized M-estimation, *Journal of Statistical Planning and Inference*, 83, 395–412.

Wiens, D. P. (2005a), Robust allocation schemes for clinical trials with prognostic factors, *Journal of Statistical Planning and Inference*, 127, 323–340.

Wiens, D. P. (2005b), Robustness in spatial studies II: Minimax design, *Environmetrics*, 16, 205–217.

Wiens, D. P. (2009), Robust discrimination designs, *Journal of the Royal Statistical Society (Series B*), 71, 805–829.

Wiens, D. P. (2010), Robustness of design for the testing of lack of fit and for estimation in binary response models, *Computational Statistics and Data Analysis*, 54, 3371–3378.

Wiens, D. P. and Wu, E. K. H. (2010), A comparative study of robust designs for M-estimated regression models, *Computational Statistics and Data Analysis*, 54, 1683–1695.

Wiens, D. P. and Xu, X. (2008a), Robust prediction and extrapolation designs for misspecified generalized linear regression models, *Journal of Statistical Planning and Inference*, 138, 30–46.

Wiens, D. P. and Xu, X. (2008b), Robust designs for one-point extrapolation, *Journal of Statistical Planning and Inference*, 138, 1339–1357.

Wiens, D. P. and Zhou, J. (1996), Minimax regression designs for approximately linear models with autocorrelated errors, *Journal of Statistical Planning and Inference*, 55, 95–106.

Wiens, D. P. and Zhou, J. (1997), Robust designs based on the infinitesimal approach, *Journal of the American Statistical Association*, 92, 1503–1511.

Wiens, D. P. and Zhou, J. (1999), Minimax designs for approximately linear models with AR(1) errors, *The Canadian Journal of Statistics*, 27, 781–794.

Wiens, D. P. and Zhou, J. (2008), Robust estimators and designs for field experiments, *Journal of Statistical Planning and Inference*, 138, 93–104.

Wierich, W. (1988), A-optimal design measures for one-way layouts with additive regression, *Journal of Statistical Planning and Inference*, 18, 57–68.

Woods, D. C., Lewis, S. M., Eccleston, J. A., and Russell, K. G. (2006), Designs for generalized linear models with several variables and model uncertainty, *Technometrics*, 48, 284–292.

Wu, C.-F. (1981), On the robustness and efficiency of some randomized designs, *Annals of Statistics*, 9, 1168–1177.

Xie, M.-Y. and Fang, K.-T. (2000), Admissibility and minimaxity of the uniform design measure in nonparametric regression model, *Journal of Statistical Planning and Inference*, 83, 101–111.

Xu, X. (2009a), Robust designs for misspecified exponential regression models, *Applied Stochastic Models in Business and Industry*, 25, 179–193.

Xu, X. (2009b), Robust prediction and extrapolation designs for censored data, *Journal of Statistical Planning and Inference*, 139, 486–502.

Xu, X. and Yuen, W. K. (2011), Applications and implementations of continuous robust designs, *Communications in Statistics – Theory and Methods*, 40, 969–988.

Yue, R.-X. and Hickernell, F. J. (1999), Robust designs for fitting linear models with misspecification, *Statistica Sinica*, 9, 1053–1069.

Zhou, J. (2001a), Integer-valued, minimax robust designs for approximately linear models with correlated errors, *Communications in Statistics: Theory and Methods*, 30, 2139.

Zhou, J. (2001b), A robust criterion for experimental designs for serially correlated observations, *Technometrics*, 43, 462–467.

Zhou, J. (2008), D-optimal minimax regression designs on discrete design space, *Journal of Statistical Planning and Inference*, 138, 4081–4092.

Zhu, W. and Wong, W.-K. (2001), Bayesian optimal designs for estimating a set of symmetrical quantiles, *Statistics in Medicine*, 20, 123–137.