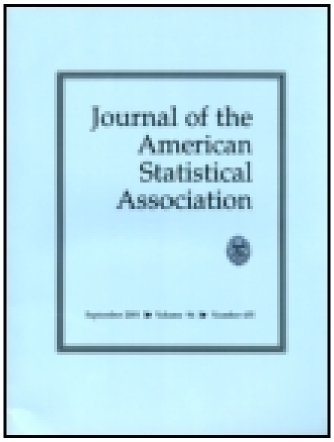

Journal of the American Statistical Association
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/uasa20

Model-Robust Designs for Quantile Regression

Linglong Kong & Douglas P. Wiens

Accepted author version posted online: 10 Oct 2014.Published online: 22 Apr 2015.

Model-Robust Designs for Quantile Regression

Linglong KONG and Douglas P. WIENS

We give methods for the construction of designs for regression models, when the purpose of the investigation is the estimation of the conditional quantile function, and the estimation method is quantile regression. The designs are robust against misspecified response functions, and against unanticipated heteroscedasticity. The methods are illustrated by example, and in a case study in which they are applied to growth charts.

KEY WORDS: Asymptotic mean squared error; B-splines; Compound design; Exchange algorithm; Genetic algorithm; Growth charts; Heteroscedasticity; Minimax bias; Minimax mean squared error; Nonlinear models; Regression quantiles; Uniformity.

1. INTRODUCTION

The need for robust methods of analysis in statistical investigations was convincingly made by Huber (1981), in whose work one finds a concentration on robustness against departures from the investigator's assumed parametric model of the *distribution* generating the data. Box and Draper (1959) had earlier made the case that, when there is any doubt about the form of the *response model* in a regression analysis in which the choice of design is under the control of the experimenter, then such choices should be made robustly, that is, with an eye to the performance of the resulting designs under a range of plausible alternate models. A focus of the work of Box and Draper was on designs robust against polynomial responses of degrees higher than that anticipated by the experimenter. This was extended in one direction by Huber (1975), who derived minimax designs for straight line fits; these minimize the maximum mean squared error of the fitted values, with the maximum taken over a full L_2-neighborhood of the experimenter's assumed response. This work, for which it was assumed that the regression estimates would be obtained by least squares, has in turn been extended in numerous directions—by Li (1984) to finite design spaces; by Wiens (1992) to multiple regression; by Woods et al. (2006) to GLMs; and by Li and Wiens (2011) to dose-response studies, to list but a few.

A method of estimation with a degree of *distributional robustness* is M-estimation (Huber 1964). Such methods convey robustness against outliers in the response variable of a regression, but have influence functions which are unbounded in the factor space. For random regressors this unboundedness may be addressed by the use of bounded influence (BI) methods (Maronna and Yohai 1981; Simpson, Ruppert, and Carroll 1992); otherwise it can be controlled by the design. Designs to be used in league with M- or BI-estimates have been studied by Wiens (2000) and Wiens and Wu (2010). In the latter article it was found that there is very little difference between designs optimal (in some sense, robust or not) for least squares and those for M-estimation; this is, however, not the case for BI-estimation.

An increasingly popular method of estimation and inference was furnished by Koenker and Bassett (1978), who elegantly restated the case for robustness, went on to extend the notion of univariate quantiles to regression quantiles, and derived *quantile regression* methods of estimating the conditional quantile function. Koenker and Bassett point out that the influence function of a quantile regression estimator is, like that of an M-estimator, unbounded in the factor space. This can again be addressed by the design. Dette and Trampisch (2012) recently studied this problem, assuming that the experimenter's assumed model is correct; to date there is no published work on designs for quantile regression methods, which extends the natural robustness of these methods against outliers to robustness against misspecified response models. We do so in this article, and also consider robustness against unanticipated heteroscedasticity.

The need for optimal designs for quantile regression methods was convincingly articulated by Dette and Trampisch (2012). That for robustness of design can arise in numerous ways. Beyond the obvious—that in many studies the fitted model is adopted largely as an article of faith—there are numerous scenarios in which the final goal is to fit models which might not fall within a standard design paradigm, but for which a preliminary study with reasonable efficiency against a range of models might furnish a point from which to expand the investigations. Some recent examples of this employ quantile regression in model selection (Behl, Claeskes, and Dette 2014), ecological studies (Martínez-Silva et al. 2013), financial modeling (Rubia and Sanchis-Marco 2013), and the fitting of time-varying coefficients (Ma and Wei 2012). In such cases the initial fitted model might be nonlinear; we address this in Section 2.2.

In Section 2 we outline our notion of misspecified response models, and set the stage for the optimality problems to be addressed in subsequent sections. The misspecification engenders a bias in the estimate, motivating our use of mean squared error (MSE) of the estimate of the conditional quantile function as a measure of the loss. In Section 3 we illustrate some designs which minimize the maximum MSE, with this maximum taken over certain very broad classes of response misspecifications. Then in Section 4 we specialize to designs which address only the bias component of the MSE—this is somewhat of a return to the findings of Box and Draper, who stated that "... the optimal design in typical situations in which both variance and bias

Linglong Kong (*lkong@ualberta.ca*) and Douglas P. Wiens (*doug.wiens@ualberta.ca*), Department of Mathematical and Statistical Sciences, University of Alberta, Edmonton, Alberta, Canada T6G 2G1. This work has been supported by the Natural Sciences and Engineering Research Council of Canada.

Color versions of one or more of the figures in the article can be found online at *www.tandfonline.com/r/jasa*.

Journal of the American Statistical Association
March 2015, Vol. 110, No. 509, Theory and Methods
DOI: 10.1080/01621459.2014.969427

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).

The bias-minimizing designs turn out to have design weights proportional to the square roots of the variance functions, when these functions are known. If instead they are also allowed to range over a certain broad class of variance functions, then *uniform* designs minimize the maximum bias over both types of departures from the experimenter's assumptions. In Section 5 this optimality of uniform designs is extended to minimization of the maximum MSE over both types of departures; we find that the minimax designs are uniform on their support. Finally, in Section 6, we illustrate the theory we have developed in an application to growth charts, in which the regressors are cubic B-splines and the appropriate choice of knots, and their locations, is in doubt.

We have posted software (see *http://www.stat.ualberta.ca/˜wiens/homepage/pubs/qrd.zip*) which runs on MATLAB, and instructions for its use, to compute the optimal designs in all of these scenarios. All derivations and longer mathematical arguments are in the Appendix, or in the online addendum Kong and Wiens (2014).

2. APPROXIMATE QUANTILE REGRESSION MODELS

To set the stage for the examples of subsequent sections, suppose that an experimenter intends to make observations on random variables Y with structure

$$Y = \boldsymbol{f}'(\boldsymbol{x})\boldsymbol{\theta} + \sigma(\boldsymbol{x})\varepsilon, \tag{1}$$

for a p-vector $\boldsymbol{f}$ of functionally independent regressors, each element of which is a function of a q-vector $\boldsymbol{x}$ of independent variables chosen (the "design") from a space χ. We assume that the errors ε are iid, and that the variance function $\sigma^2(\boldsymbol{x})$ is strictly positive on the support of the design. For a fixed $\tau \in (0,1)$, $\boldsymbol{f}'(\boldsymbol{x})\boldsymbol{\theta}$ is to be the conditional τ-quantile of Y, given $\boldsymbol{x}$:

$$\tau = G_\varepsilon(0) = G_{Y|\boldsymbol{x}}(\boldsymbol{f}'(\boldsymbol{x})\boldsymbol{\theta}). \tag{2}$$

(We write $G_U(\cdot)$ for the distribution function of a random variable U.)

Now suppose that (1) is only an approximation, and that in fact

$$Y = \boldsymbol{f}'(\boldsymbol{x})\boldsymbol{\theta} + \delta_n(\boldsymbol{x}) + \sigma(\boldsymbol{x})\varepsilon, \tag{3}$$

for some "small" model error δ_n. The dependence of δ on n is necessary for a sensible asymptotic treatment—in order that bias and variance remain of the same order we will assume that $\delta_n = O(n^{-1/2})$. For fixed sample sizes this is not necessary.

The experimenter, acting as though $\delta_n \equiv 0$ and $\sigma(\cdot)$ is constant, computes the quantile regression estimate

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{t}} \sum_{i=1}^{n} \rho_\tau(Y_i - \boldsymbol{f}'(\boldsymbol{x}_i)\boldsymbol{t}), \tag{4}$$

where $\rho_\tau(\cdot)$ is the "check" function $\rho_\tau(r) = r(\tau - I(r<0))$, with derivative $\psi_\tau(r) = \tau - I(r<0)$.

We will consider two types of design spaces χ. The first is discrete, with N possible design points $\{\boldsymbol{x}_i\}_{i=1}^N$; here N is arbitrary. We also consider a continuous, compact design space, with Lebesgue measure $\mathrm{VOL}(\chi) \stackrel{\text{def}}{=} \int_\chi d\boldsymbol{x}$, in which case the design is generated by a design measure $\xi(d\boldsymbol{x})$. Initially, we shall unify the presentation by writing sums of the form $\sum_{\boldsymbol{x}\in\text{design}} \alpha(\boldsymbol{x})$, in which a fraction $\xi_{n,i} = n_i/n$ of the n observations are to be made at the design point $\boldsymbol{x} = \boldsymbol{x}_i$, as Lebesgue-Stieltjes integrals, viz, as $n\sum_{i=1}^N \xi_{n,i}\alpha(\boldsymbol{x}_i) = n\int_\chi \alpha(\boldsymbol{x})\xi_n(d\boldsymbol{x})$. We assume that the design measure ξ_n has a weak limit ξ_∞ for which

$$\lim_{n\to\infty} \sum_{i=1}^{n} \xi_{n,i}\rho_\tau(Y_i - \boldsymbol{f}'(\boldsymbol{x}_i)\boldsymbol{t}) = \int_\chi E_{Y|\boldsymbol{x}}[\rho_\tau(Y - \boldsymbol{f}'(\boldsymbol{x})\boldsymbol{t})]\xi_\infty(d\boldsymbol{x}).$$

Under (3) the meaning of $\boldsymbol{\theta}$ becomes ambiguous. Thus we define this "true" regression parameter as that making the experimenter's approximation (1) most accurate, under the experimenter's assumption of homoscedasticity. For a discrete design space this is

$$\boldsymbol{\theta} = \arg\min_{\boldsymbol{t}} \frac{1}{N}\sum_{i=1}^{N} E_{Y|\boldsymbol{x}}[\rho_\tau(Y - \boldsymbol{f}'(\boldsymbol{x}_i)\boldsymbol{t})]. \tag{5}$$

Carrying out the minimization in (5) and evaluating at $\boldsymbol{t} = \boldsymbol{\theta}$:

$$\begin{aligned}\boldsymbol{0} &= \frac{1}{N}\sum_{i=1}^{N} E_{Y|\boldsymbol{x}}[\psi_\tau(Y - \boldsymbol{f}'(\boldsymbol{x}_i)\boldsymbol{t})]\boldsymbol{f}(\boldsymbol{x}_i)\\ &= \frac{1}{N}\sum_{i=1}^{N}[G_\varepsilon(0) - G_\varepsilon(-\delta_n(\boldsymbol{x}_i))]\boldsymbol{f}(\boldsymbol{x}_i)\\ &= (g_\varepsilon(0) + O(n^{-1/2}))\frac{1}{N}\sum_{i=1}^{N}\delta_n(\boldsymbol{x}_i)\boldsymbol{f}(\boldsymbol{x}_i),\end{aligned} \tag{6}$$

where g_ε is the density of G_ε. We now define $\delta_0(\boldsymbol{x}) = \lim_{n\to\infty}\sqrt{n}\delta_n(\boldsymbol{x})$, so that

$$\frac{1}{N}\sum_{i=1}^{N}\delta_0(\boldsymbol{x}_i)\boldsymbol{f}(\boldsymbol{x}_i) = \boldsymbol{0}. \tag{7}$$

In a continuous design space the average is replaced by an integral—see Equation (8b).

The true conditional τ-quantile $Y_\tau = \boldsymbol{f}'(\boldsymbol{x})\boldsymbol{\theta} + \delta_n(\boldsymbol{x})$ is predicted by $\hat{Y}_\tau = \boldsymbol{f}'(\boldsymbol{x})\hat{\boldsymbol{\theta}}$, and our approach is to obtain the asymptotic mean squared error matrix $\mathrm{MSE}_{\hat{\boldsymbol{\theta}}}$ of the parameter estimates, thus obtaining the average—over χ—MSE of these predicted values, and to maximize this average MSE over the appropriate choice

$$\chi \text{ discrete: } \Delta_0 = \left\{\delta_0(\cdot)\,|\ \text{(i) } N^{-1}\sum_{i=1}^{N}\delta_0(\boldsymbol{x}_i)\boldsymbol{f}(\boldsymbol{x}_i) = \boldsymbol{0} \text{ and (ii) } N^{-1}\sum_{i=1}^{N}\delta_0^2(\boldsymbol{x}_i) \le \eta^2\right\}, \tag{8a}$$

$$\chi \text{ continuous: } \Delta_0 = \left\{\delta_0(\cdot)\,|\ \text{(i) } \int_\chi \delta_0(\boldsymbol{x})\boldsymbol{f}(\boldsymbol{x})d\boldsymbol{x} = \boldsymbol{0} \text{ and (ii) } \int_\chi \delta_0^2(\boldsymbol{x})d\boldsymbol{x} \le \eta^2\right\}. \tag{8b}$$

This is carried out in Section 2.3. We also consider classes of variance functions. These may be independent of the design or—see Equation (18)—vary with the designs weights, in which case we also maximize the MSE over this class. In any event we then go on to find the MSE-minimizing designs ξ_*, using a variety of analytic and numerical techniques.

In most cases the optimal designs ξ_* must be approximated in order to implement them in finite samples; for example, when $q = 1$ we will do this by placing the design points at the quantiles

$$x_i = \xi_*^{-1}\left(\frac{i-0.5}{n}\right), \tag{9}$$

or at the closest available points in discrete design spaces. For $q > 1$ the situation is more interesting and some suggestions are in Fang and Wang (1994) and Xu and Yuen (2011); an intriguing possibility as yet (to our knowledge) unexplored is the use of vector quantization to approximate the designs.

2.1 Asymptotics

In Equation (8) the imposition of Equation (7), and its analogue in continuous spaces, ensures the identifiability of the parameter in (3). The bounds of η^2 force the errors due to variation, and those due to the bias engendered by the model misspecification, to remain of the same order asymptotically—a situation akin to the imposition of contiguity in the asymptotic theory of hypothesis testing. Define

$$\boldsymbol{\mu}_0 = \int_\chi \delta_0(\boldsymbol{x})\frac{1}{\sigma(\boldsymbol{x})}\boldsymbol{f}(\boldsymbol{x})\xi_\infty(d\boldsymbol{x}), \tag{10a}$$

$$\boldsymbol{P}_0 = \int_\chi \boldsymbol{f}(\boldsymbol{x})\boldsymbol{f}'(\boldsymbol{x})\xi_\infty(d\boldsymbol{x}), \tag{10b}$$

$$\boldsymbol{P}_1 = \int_\chi \boldsymbol{f}(\boldsymbol{x})\frac{1}{\sigma(\boldsymbol{x})}\boldsymbol{f}'(\boldsymbol{x})\xi_\infty(d\boldsymbol{x}). \tag{10c}$$

Assume that the support of ξ_∞ is large enough that $\boldsymbol{P}_0$ and $\boldsymbol{P}_1$ are positive definite. Define the target parameter $\boldsymbol{\theta}$ to be the asymptotic solution to (4), so that

$$\sum_{i=1}^n \xi_{n,i}\psi_\tau(Y_i - \boldsymbol{f}'(\boldsymbol{x}_i)\boldsymbol{\theta})\boldsymbol{f}(\boldsymbol{x}_i) \xrightarrow{\text{pr}} \boldsymbol{0}, \tag{11}$$

in agreement with (6). The proof of the asymptotic normality of the estimate runs along familiar lines—see Knight (1998) and Koenker (2005)—and so we merely state the result. Complete details are in Kong and Wiens (2014).

Theorem 1. Under conditions (A1)–(A3) the quantile regression estimate $\hat{\boldsymbol{\theta}}_n$ of the parameter $\boldsymbol{\theta}$ defined by Equation (11) is asymptotically normally distributed:

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}) \xrightarrow{L} N\left(\boldsymbol{P}_1^{-1}\boldsymbol{\mu}_0, \frac{\tau(1-\tau)}{g_\varepsilon^2(0)}\boldsymbol{P}_1^{-1}\boldsymbol{P}_0\boldsymbol{P}_1^{-1}\right).$$

2.2 Nonlinear Models

Dette and Trampisch (2012) obtained (nonrobust) designs for quantile regression and *nonlinear* models; these were *locally optimal*, with, in our notation, (2) replaced by $\tau = G_{Y|\boldsymbol{x}}(F(\boldsymbol{x};\boldsymbol{\theta}))$, where $F(\boldsymbol{x};\boldsymbol{\theta})$ is evaluated at fixed values of those elements of $\boldsymbol{\theta}$ which enter in a nonlinear manner. Some robustness against misspecifications of these local parameters was then introduced by considering Bayesian and maximin designs. In this article the examples pertain only to *linear* models. However, since (as also in Dette and Trampisch 2012), our approach is asymptotic in nature, the results presented here are easily modified to accommodate nonlinear models. The definition (4) of the estimate is replaced by $\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{t}} \sum_{i=1}^n \rho_\tau(Y_i - F(\boldsymbol{x}_i;\boldsymbol{t}))$, and then, in all occurrences, $\boldsymbol{f}(\boldsymbol{x})$ is to be replaced by the gradient $\boldsymbol{f}_{\boldsymbol{\theta}}(\boldsymbol{x}) = \partial F(\boldsymbol{x};\boldsymbol{\theta})/\partial\boldsymbol{\theta}$. With these changes Theorem 1 continues to hold, as does the rest of the theory of the article. The robustness is then attained against misspecifications in the functional form of $F(\boldsymbol{x};\cdot)$, possibly but not necessarily arising from misspecified parameters.

2.3 Maximum MSE Over Δ_0; Discrete Design Spaces

From Theorem 1, the asymptotic MSE matrix of $\hat{\boldsymbol{\theta}}$ is

$$\text{MSE}_{\hat{\boldsymbol{\theta}}} = \boldsymbol{P}_1^{-1}\left[\frac{\tau(1-\tau)}{g_\varepsilon^2(0)}\boldsymbol{P}_0 + \boldsymbol{\mu}_0\boldsymbol{\mu}_0'\right]\boldsymbol{P}_1^{-1}.$$

We now introduce a measure of the asymptotic loss when the conditional quantile $Y_\tau(\boldsymbol{x}) = \boldsymbol{f}'(\boldsymbol{x})\boldsymbol{\theta} + \delta_n(\boldsymbol{x})$, for $\boldsymbol{x} \in \chi$, is incorrectly estimated by $\hat{Y}_n(\boldsymbol{x}) = \boldsymbol{f}'(\boldsymbol{x})\hat{\boldsymbol{\theta}}_n$. For a discrete design space $\chi = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N\}$ this measure is the limiting average mean squared error

$$\text{AMSE} = \lim_n \frac{1}{N}\sum_{i=1}^N E[\{\sqrt{n}(\hat{Y}_n(\boldsymbol{x}_i) - Y_\tau(\boldsymbol{x}_i))\}^2].$$

In terms of $\boldsymbol{A} = N^{-1}\sum_{i=1}^N \boldsymbol{f}(\boldsymbol{x}_i)\boldsymbol{f}'(\boldsymbol{x}_i)$, and using (i) of (8a), we find that

$$\begin{aligned}\text{AMSE} &= \text{tr}\left(\boldsymbol{A}\cdot\text{MSE}_{\hat{\boldsymbol{\theta}}}\right) + \frac{1}{N}\sum_{i=1}^N \delta_0^2(\boldsymbol{x}_i)\\ &= \frac{\tau(1-\tau)}{g_\varepsilon^2(0)}\text{tr}\left(\boldsymbol{A}\boldsymbol{P}_1^{-1}\boldsymbol{P}_0\boldsymbol{P}_1^{-1}\right) + \boldsymbol{\mu}_0'\boldsymbol{P}_1^{-1}\boldsymbol{A}\boldsymbol{P}_1^{-1}\boldsymbol{\mu}_0\\ &\quad + \frac{1}{N}\sum_{i=1}^N \delta_0^2(\boldsymbol{x}_i).\end{aligned} \tag{12}$$

We now write merely ξ for ξ_∞. We impose a bound $N^{-1}\sum_{i=1}^N \sigma^2(\boldsymbol{x}_i) \le \sigma_0^2$ for a given σ_0^2, and we denote by $ch_{\max}$ the maximum eigenvalue of a matrix. The maximum value of AMSE over Δ_0 is given in the following theorem.

Theorem 2. For a discrete design space χ define

$$\boldsymbol{T}_{0,0} = \sum_{\xi_i>0}\boldsymbol{f}(\boldsymbol{x}_i)\boldsymbol{f}'(\boldsymbol{x}_i)\xi_i,\ \boldsymbol{T}_{0,k} = \sum_{\xi_i>0}\boldsymbol{f}(\boldsymbol{x}_i)\boldsymbol{f}'(\boldsymbol{x}_i)\left(\frac{\xi_i}{\sigma(\boldsymbol{x}_i)/\sigma_0}\right)^k,\ k = 1, 2,$$

and

$$\boldsymbol{T}_0 = \boldsymbol{T}_{0,1}^{-1}\boldsymbol{T}_{0,0}\boldsymbol{T}_{0,1}^{-1},\ \boldsymbol{T}_2 = \boldsymbol{T}_{0,1}^{-1}\boldsymbol{T}_{0,2}\boldsymbol{T}_{0,1}^{-1}. \tag{13}$$

Then $\max_{\Delta_0}$AMSE is $\frac{\tau(1-\tau)\sigma_0^2}{g_\varepsilon^2(0)} + \eta^2$ times

$$\mathcal{L}_\nu(\xi|\sigma) = (1-\nu)\,\text{tr}(\boldsymbol{A}\boldsymbol{T}_0) + \nu ch_{\max}(\boldsymbol{A}\boldsymbol{T}_2), \tag{14}$$

where $\nu = \eta^2/\{\frac{\tau(1-\tau)\sigma_0^2}{g_\varepsilon^2(0)} + \eta^2\}$.

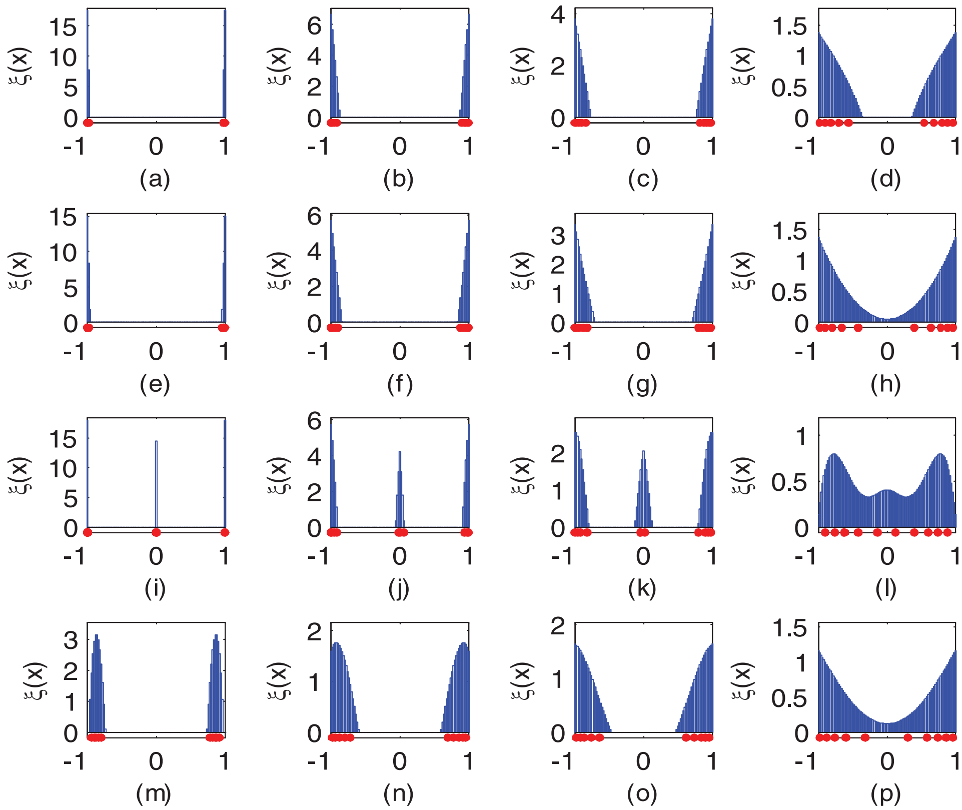

Figure 1. Minimax (over Δ_0) design measures for heteroscedastic straight line regression, $N = 101$, normalized so that bar area = 1. Columns 1–4 use $\nu = 0.05, 0.35, 0.65, 0.95$, respectively; rows 1–4 use $\sigma(x) \propto (1 + |x|)^{-1}$, 1, $0.2 + |x|$, $1 + (x/2)^2$, respectively. The bullets below the horizontal axes are the locations of $n = 10$ design points, implemented as at (9).

The first component ($\mathrm{tr}(\boldsymbol{AT}_0)$) of $\mathcal{L}_\nu(\xi|\sigma)$ arises solely from variation, the second ($ch_{\max}(\boldsymbol{AT}_2)$) from (squared) bias. Note that (14) depends on σ_0 only through $\{\sigma(\boldsymbol{x})/\sigma_0\}$ and through ν. We may thus without loss of generality take $\sigma_0 = 1$ and parameterize the designs solely by $\nu \in [0, 1]$, which may be chosen by the experimenter, representing his relative concern for errors due to bias rather than to variation.

3. EXAMPLES: DESIGNS MINIMIZING MAX$_{\Delta_0}$ MSE FOR FIXED VARIANCE FUNCTIONS

Before extending the theory presented thus far, we illustrate it for some representative, fixed variance functions in two cases—approximate straight line regression in a discrete design space, and approximate quadratic regression in a continuous design space. The development of the first case is given in some detail in the Appendix; that for the second is outlined only briefly.

3.1 Discrete Design Spaces

For least squares regression problems with univariate design variables and homoscedastic variances, optimally robust designs have been constructed by, among others, Fang and Wiens (2000), who computed exact designs by simulated annealing. Here we construct optimal designs for heteroscedastic quantile regression problems and also take a different approach to the implementation—we obtain exact optimal values $\{\xi_{*,i}\}$ and then implement the designs as at (9).

For a fixed variance function and a discrete design space we seek a design ξ_* minimizing (14). We illustrate the method in the case of approximate straight line models—$\boldsymbol{f}(x_i) = (1, x_i)'$—and suppose that the design space χ consists of N points in $[-1, 1]$. The space χ is symmetric in that if $\chi = (x_1, \ldots, x_N)'$ $(-1 = x_1 < \cdots < x_N = 1)$ and χ_π denotes the reversal $(x_N, \ldots, x_1)'$ then $\chi_\pi = -\chi$. We consider symmetric designs, that is, designs for which $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_N)'$, with $\xi_i = \xi(x_i)$, satisfies $\xi(x_i) = \xi(-x_i)$. We also assume a symmetric but arbitrary variance function $\sigma_i = \sigma(|x_i|)$.

The designs are obtained by variational arguments followed by a constrained numerical minimization; the details are in the Appendix. See Figure 1 for representative plots of the designs, scaled so as to have unit area. In these plots the bullets below the horizontal axes are the locations of $n = 10$ design points, implemented as at (9). In the case of homoscedasticity (plots (e)–(h)) the designs for very small ν are close in nature to their nonrobust counterparts, placing point masses at ± 1. As ν increases these replicates spread out into clusters near ± 1 and, depending upon the variance function, possibly near 0 as well. The limiting behavior as $\nu \to 1$ is studied in Section 4.

3.2 Continuous Design Spaces

The continuous case requires special consideration. Rather than AMSE at (12) we use instead the integrated mean squared

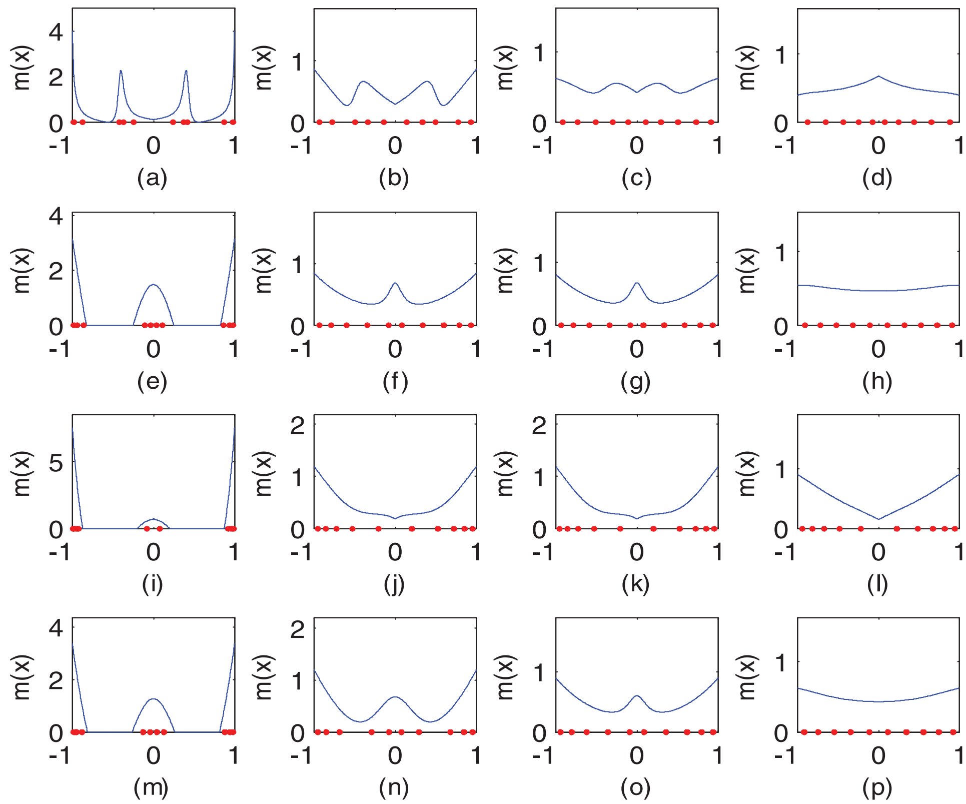

Figure 2. Minimax (over Δ_0) design densities for heteroscedastic quadratic regression on $[-1, 1]$. Columns 1–4 use $\nu = 0.05, 0.35, 0.65, 0.95$, respectively; rows 1–4 use $\sigma(x) \propto (1 + |x|)^{-1}$, 1, $0.2 + |x|$, $1 + (x/2)^2$, respectively. The bullets on the horizontal axes are the locations of $n = 10$ design points, implemented as in (9).

error

$$\text{IMSE} = \lim_n \int_\chi E[\{\sqrt{n}(\hat{Y}_n(\boldsymbol{x}) - Y_\tau(\boldsymbol{x}))\}^2] d\boldsymbol{x},$$

together with $\boldsymbol{A} = \int_\chi \boldsymbol{f}(\boldsymbol{x})\boldsymbol{f}'(\boldsymbol{x}) d\boldsymbol{x}$, and obtain

$$\text{IMSE} = \frac{\tau(1-\tau)}{g_\varepsilon^2(0)} \text{tr}\left(\boldsymbol{A}\boldsymbol{P}_1^{-1}\boldsymbol{P}_0\boldsymbol{P}_1^{-1}\right) + \boldsymbol{\mu}_0'\boldsymbol{P}_1^{-1}\boldsymbol{A}\boldsymbol{P}_1^{-1}\boldsymbol{\mu}_0 + \int_\chi \delta_0^2(\boldsymbol{x})\, d\boldsymbol{x}.$$

In order that the maximum IMSE be finite, it is necessary that the design measure be absolutely continuous. That this should be so is intuitively clear—if ξ_∞ in (10) places positive mass on sets of Lebesgue measure zero, such as individual points, then δ_0 may be chosen arbitrarily large on such sets without altering its membership in Δ_0, and one can do this in such a way as to drive IMSE beyond all bounds, through (10a). A formal proof may be based on that of Lemma 1 in Heo, Schmuland, and Wiens (2001).

When implementing continuous designs we discretize; for instance, when there is only one covariate we employ (9). As a referee has pointed out, this might result in an unbounded IMSE along particularly pathological sequences $\{\delta_n\}$. A possible alternative, which we do not illustrate here since it is unlikely to find favor with practitioners, is to randomly choose design points from the optimal design measure; in the parlance of game theory this would thwart the intentions of a malevolent Nature, which can then not anticipate the design.

In the same vein Bischoff (2010) stated a criticism, in a context of discretized, absolutely continuous, lack-of-fit designs as proposed by Wiens (1991) and Biedermann and Dette (2001), of the very rich class of alternatives, analogous to (8b), used by those authors. Bischoff suggested using a smaller class of alternatives; here we are, however, in accord with Wiens (1992), who states, "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."

We write $m(\boldsymbol{x})$ for the density of ξ when dealing with continuous design spaces and take $\int_\chi \sigma^2(\boldsymbol{x}) d\boldsymbol{x} \le \sigma_0^2$ (= 1, as in the discrete case).

Theorem 3. For a continuous design space χ define $\boldsymbol{T}_0$ and $\boldsymbol{T}_2$ as at (13), with

$$\boldsymbol{T}_{0,0} = \int_\chi \boldsymbol{f}(\boldsymbol{x})\boldsymbol{f}'(\boldsymbol{x}) m(\boldsymbol{x})\, d\boldsymbol{x} \text{ and}$$

$$\boldsymbol{T}_{0,k} = \int_\chi \boldsymbol{f}(\boldsymbol{x})\boldsymbol{f}'(\boldsymbol{x}) \left(\frac{m(\boldsymbol{x})}{\sigma(\boldsymbol{x})/\sigma_0}\right)^k d\boldsymbol{x},\ k = 1, 2.$$

Then the maximum IMSE is given by Equation (14).

As an example we minimize IMSE for approximate quadratic regression, that is, $\boldsymbol{f}(x) = (1, x, x^2)'$, and a fixed variance function $\sigma^2(x)$, over the design space $\chi = [-1, 1]$. Similar problems, assuming homoscedasticity, were studied previously by Shi, Ye, and Zhou (2003) using methods of nonsmooth optimization, and by Daemi and Wiens (2013) following the methods used here and outlined in the Appendix.

We show in the Appendix that the minimizing density is of the form

$$m(x;\mathbf{a}) = \left(\frac{q_1(x)\sigma(x) + q_2(x)}{a_{00} + \frac{q_3(x)}{\sigma(x)}}\right)^+, \quad (15)$$

for polynomials $q_j(x) = a_{0j} + a_{2j}x^2 + a_{4j}x^4$, $j = 1, 2, 3$. The 10 constants a_{ij} forming $\mathbf{a}$ are chosen to minimize the loss $\mathcal{L}_\nu(\xi|\sigma)$ at (14) over $\mathbf{a}$, subject to $\int_{-1}^{1} m(x;\mathbf{a})dx = 1$. Some examples are illustrated in Figure 2. Again there is a pronounced increase in the spreading out of the mass as ν increases, and again under homoscedasticity these masses are initially concentrated near ± 1 and 0, as in the nonrobust case. It is rather evident from the plots in the rightmost panels of Figure 2 that, as $\nu \to 1$, the density $m(x;\mathbf{a})$ becomes proportional to $\sigma(x)$, a phenomenon explained in the following section.

4. BIAS MINIMIZING DESIGNS

The following result is quite elementary, but since we use it repeatedly we give it a formal statement and proof.

Proposition 1. (i) Suppose that χ is discrete, that the function $p(\boldsymbol{x})$ is defined on $\chi_0 \subset \chi$ and that $\mathbf{M}_q \stackrel{\text{def}}{=} \sum_{\boldsymbol{x}_i \in \chi_0} q(\boldsymbol{x}_i)\boldsymbol{f}(\boldsymbol{x}_i)\boldsymbol{f}'(\boldsymbol{x}_i)$ exists for $q = p$, $q = p^2$ and $q = \mathbf{1}$ ($\mathbf{1}(\boldsymbol{x}_i) \equiv 1$), and is invertible for $q = p$ and $q = \mathbf{1}$. Then, under the ordering "$\succeq$" with respect to positive semidefiniteness,

$$\mathbf{M}_p^{-1}\mathbf{M}_{p^2}\mathbf{M}_p^{-1} \succeq \mathbf{M}_{\mathbf{1}}^{-1}. \quad (16)$$

(ii) Suppose that χ is continuous, that the function $p(\boldsymbol{x})$ is defined on $\chi_0 \subset \chi$ and that $\mathbf{M}_q \stackrel{\text{def}}{=} \int_{\chi_0} \boldsymbol{f}(\boldsymbol{x})\boldsymbol{f}'(\boldsymbol{x})q(\boldsymbol{x})\,d\boldsymbol{x}$ exists for $q = p$, $q = p^2$ and $q = \mathbf{1}$, and is invertible for $q = p$ and $q = \mathbf{1}$. Then (16) holds.

In discrete design spaces we define $\boldsymbol{A}_\xi = \sum_{\xi_i > 0} \boldsymbol{f}(\boldsymbol{x}_i)\boldsymbol{f}'(\boldsymbol{x}_i)$. It then follows from Proposition 1 that $\boldsymbol{T}_2 \succeq \boldsymbol{A}_\xi^{-1}$; note as well that $\boldsymbol{A}_\xi^{-1} \succeq (N\boldsymbol{A})^{-1}$. Together these imply that

$$\mathcal{L}_{\nu=1}(\xi|\sigma) = ch_{\max}(\mathbf{A}\boldsymbol{T}_2) \geq ch_{\max}\left(\mathbf{A}\mathbf{A}_\xi^{-1}\right) \geq 1/N. \quad (17)$$

Motivated by the remark of Box and Draper (1959) quoted in Section 1 of this article we note that, if the experimenter seeks robustness only against errors due to bias (so that $\nu = 1$), whether arising from a misspecified response model or a particular variance function $\sigma^2(\cdot)$, then the maximum bias is minimized by $\xi_i = \sigma(\boldsymbol{x}_i)/\sum_{i=1}^{N}\sigma(\boldsymbol{x}_i)$, since then $\boldsymbol{T}_2 = \mathbf{A}_\xi^{-1}$ and the lower bound in (17) is attained.

Similarly, in a continuous design space the maximum bias is minimized by $m(\boldsymbol{x}) = \sigma(\boldsymbol{x})/\int_\chi \sigma(\boldsymbol{x})d\boldsymbol{x}$; for this we use $\boldsymbol{A}_m = \int_{m(\boldsymbol{x})>0} \boldsymbol{f}(\boldsymbol{x})\boldsymbol{f}'(\boldsymbol{x})d\boldsymbol{x}$, in place of $\boldsymbol{A}_\xi$ and obtain a lower bound of 1 in (17).

If the form of the variance function is in doubt, then a minimax approach dictates taking a further maximum over a class of such functions. We consider the class $\Sigma_0 = \left\{\sigma_\xi^2(\cdot|r) | r \in (-\infty, \infty)\right\}$ of variance functions given by

$$\sigma_\xi(\boldsymbol{x}|r) = \begin{cases} c_r \xi^{r/2}(\boldsymbol{x})\, I(\xi(\boldsymbol{x}) > 0), & \chi \text{ discrete}, \\ c_r m^{r/2}(\boldsymbol{x})\, I(m(\boldsymbol{x}) > 0), & \chi \text{ continuous}; \end{cases} \quad (18)$$

c_r is the required constant of proportionality determined by, for example, $N^{-1}\sum_{\xi_i>0}\sigma_\xi^2(\boldsymbol{x}_i|r) = 1$. In the discrete case define

$$\boldsymbol{S}_0 = \sum_{\xi_i>0} \boldsymbol{f}(\boldsymbol{x}_i)\boldsymbol{f}'(\boldsymbol{x}_i)\xi_i$$

and

$$\boldsymbol{S}_k = \boldsymbol{S}_k(r) = \sum_{\xi_i>0} \boldsymbol{f}(\boldsymbol{x}_i)\boldsymbol{f}'(\boldsymbol{x}_i)\xi_i^{k(1-\frac{r}{2})} \text{ for } k = 1, 2.$$

Note that $\boldsymbol{S}_0 = \boldsymbol{S}_1(0) = \boldsymbol{S}_2(1)$. If $\sigma^2(\cdot) \in \Sigma_0$ then $\mathcal{L}_{\nu=1}(\xi|\sigma) = ch_{\max}\left(\boldsymbol{A}\boldsymbol{S}_1^{-1}(r)\boldsymbol{S}_2(r)\boldsymbol{S}_1^{-1}(r)\right)$, which by Proposition 1 exceeds $ch_{\max}(\boldsymbol{A}\boldsymbol{A}_\xi^{-1})$; this in turn is minimized by the uniform design ξ_*, with $\xi_{*,i} \equiv 1/N$. Thus this approximate design, which we implement as at (9), is minimax with respect to bias, over Σ_0. Similarly, in a continuous design space the minimax bias design is the continuous uniform: $m_*(\boldsymbol{x}) \equiv 1/\text{VOL}(\chi)$.

This discussion has revealed why the minimax designs exhibited in Figures 1 and 2 become proportional to $\sigma(x)$ as $\nu \to 1$, and for $\nu = 1$ are uniform if the maximization is also carried out over Σ_0. For $\nu < 1$, if the form of $\sigma(\cdot)$ is known then this knowledge can be used to increase the efficiency of the design, relative to uniformity. In the next section we show that, if $\sigma(\cdot)$ is unknown but is allowed to range over Σ_0, then uniform (on their support) designs are again minimax with respect to MSE.

5. MSE MINIMIZING DESIGNS

Under (18) the maximized, over (8a), loss (14) is

$$\mathcal{L}_\nu(\xi|r) = (1-\nu)\,c_r^2 \text{tr}\left(\boldsymbol{A}\boldsymbol{S}_1^{-1}(r)\boldsymbol{S}_0\boldsymbol{S}_1^{-1}(r)\right) + \nu ch_{\max}\left(\boldsymbol{A}\boldsymbol{S}_1^{-1}(r)\boldsymbol{S}_2(r)\boldsymbol{S}_1^{-1}(r)\right).$$

Several cases are of interest for fixed r. The case $r = 0$ corresponds to homoscedasticity. That for $r = 2$ is treated in Kong and Wiens (2014). If $r = 1$ (a case which turns out to be least favorable—see the proof of Lemma 1), then

$$\mathcal{L}_\nu(\xi|r=1) = (1-\nu)N\text{tr}\left(\boldsymbol{A}\boldsymbol{S}_1^{-1}(1)\boldsymbol{S}_0\boldsymbol{S}_1^{-1}(1)\right) + \nu ch_{\max}\left(\boldsymbol{A}\boldsymbol{S}_1^{-1}(1)\boldsymbol{S}_0\boldsymbol{S}_1^{-1}(1)\right).$$

In this case the optimal, approximate design ξ_* is again uniform on all of χ: $\xi_{*,i} \equiv 1/N$. And again in the parlance of game theory, the experimenter's optimal reply to Nature's strategy of placing the variances proportional to the design weights is to design in such a way that this variance structure is in fact homoscedastic.

To see that ξ_* is uniform, note that at this design we have $\boldsymbol{S}_0 = \boldsymbol{S}_2(1) = \boldsymbol{A}$ and $\boldsymbol{S}_1(1) = \sqrt{N}\boldsymbol{A}$, so that $\boldsymbol{S}_1^{-1}(1)\boldsymbol{S}_0\boldsymbol{S}_1^{-1}(1) = (N\boldsymbol{A})^{-1}$, and it suffices to note that for any other design ξ, by Proposition 1, $\boldsymbol{S}_1^{-1}(1)\boldsymbol{S}_0\boldsymbol{S}_1^{-1}(1) \succeq \boldsymbol{A}_\xi^{-1} \succeq (N\boldsymbol{A})^{-1}$.

Similarly, in the continuous case the uniform design, with density $m_*(\boldsymbol{x})$, minimizes $\mathcal{L}_\nu(\xi|r=1)$.

To extend these optimality properties of uniform designs to all of Σ_0, we first consider the discrete case, and define $\mathcal{L}_\nu(\xi) = \max_r \mathcal{L}_\nu(\xi|r)$. By the following lemma, a minimax design is necessarily uniform on its support.

Lemma 1. If ξ is a design with k-point support $\{\boldsymbol{x}_{i_1}, \ldots, \boldsymbol{x}_{i_k}\} \subset \chi$ ($k \leq N$), placing mass ξ_{i_j} at $\boldsymbol{x}_{i_j}$, and ξ_k is the design placing mass $1/k$ at each point $\boldsymbol{x}_{i_j}$, then $\boldsymbol{A}_\xi =$

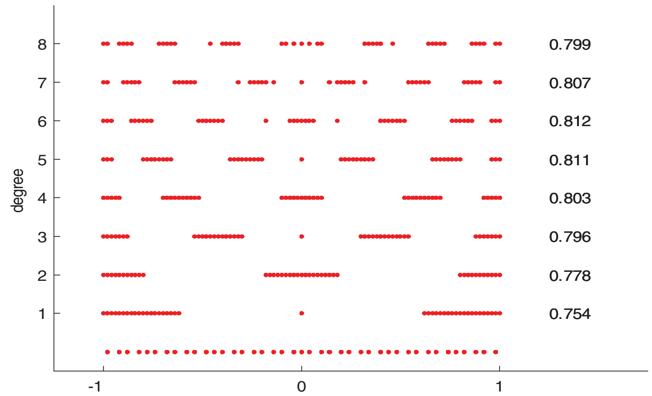

Figure 3. Minimax compound, uniform designs ξ_p minimizing $\max_{\Delta_0,\Sigma_0}$ AMSE for polynomial regression of degrees $p = 1, \ldots, 8$; $n = 41$, $N = 101$, $\nu = 0.5$. Bullets indicate design points; bottom line is the n-point implementation of the design $\xi_{*,i} \equiv 1/N$. Efficiencies $\mathcal{L}_\nu\left(\xi_p\right)/\mathcal{L}_\nu(\xi_*)$ are given at the right.

$\sum_{j=1}^{k} \boldsymbol{f}(\boldsymbol{x}_{i_j})\boldsymbol{f}'(\boldsymbol{x}_{i_j})$ and

$$\mathcal{L}_\nu(\xi) \geq \mathcal{L}_\nu(\xi_k) = (1-\nu)\, N \mathrm{tr}\left(\boldsymbol{A}\boldsymbol{A}_\xi^{-1}\right) + \nu ch_{\max}\left(\boldsymbol{A}\boldsymbol{A}_\xi^{-1}\right).$$

By Lemma 1 the search for minimax designs reduces to searching for support points on which the design is to be uniform. Since $\boldsymbol{A}_\xi$ increases, in the sense of positive semidefiniteness, as k increases, a minimax design ξ_* necessarily has maximum support size. Among approximate designs the optimal choice is thus $\xi_{*,i} \equiv 1/N$, $i = 1, \ldots, N$. Among exact designs ξ_* must have support size $k_* = \min(n, N)$; the support points $\{\boldsymbol{x}_{i_1}^*, \ldots, \boldsymbol{x}_{i_{k_*}}^*\}$ are those minimizing

$$\mathcal{L}_\nu(\xi_*) = (1-\nu)\, N \mathrm{tr}\left(\boldsymbol{A}\boldsymbol{A}_{k_*}^{-1}\right) + \nu ch_{\max}\left(\boldsymbol{A}\boldsymbol{A}_{k_*}^{-1}\right), \qquad (19)$$

with $\boldsymbol{A}_{k_*} = \sum_{j=1}^{k_*} \boldsymbol{f}(\boldsymbol{x}_{i_j}^*)\boldsymbol{f}'(\boldsymbol{x}_{i_j}^*)$. We are then seeking a compound optimal design, for which problems some general theory has been furnished by Cook and Wong (1994); in our case there is, however, the additional restriction to uniformity.

Example 5.1. In the case of straight line models and symmetric designs on a symmetric interval, $\boldsymbol{A}_{k_*} = \mathrm{diag}(k_*, \sum_{j=1}^{k_*} \boldsymbol{x}_{i_j}^{*2})$. Both components of $\mathcal{L}_\nu(\xi_*)$ are decreased by progressively including in the support the largest remaining design points, so as to "increase" $\boldsymbol{A}_{k_*}$. If n is odd, then 0 must be in the support; the remaining points—all points if n is even—are the $2 \times \min([n/2], [N/2])$ symmetrically placed design points of largest absolute value. If n is a multiple of N, say $n = mN$, then this design is replicated m times. If $n = mN + t$ for $0 < t < N$ then an exact uniform design is not attainable if $m > 0$. A possible implementation is to place m observations at each of the N points in the design space, and to append to this the $2[t/2]$ symmetrically placed design points of largest absolute value (and 0, if t is odd).

Example 5.2. We have found an exchange algorithm to be very effective at constructing compound designs minimizing (19). This has been carried out for polynomial regression over $[-1, 1]$, with the restriction to symmetric designs. See Figure 3, in which some typical cases are displayed and compared with the approximate design $\xi_{*,i} \equiv 1/N$, implemented as at (9). The efficiencies given in the figure have been found to be quite stable over other choices of n, N, and ν.

In a completely analogous manner we find that the continuous uniform design, with density $m_*(\boldsymbol{x})$, minimizes the maximum of $\mathcal{L}_\nu(\xi|r)$ over Δ_0 and Σ_0.

6. CASE STUDY: ROBUST DESIGN IN GROWTH CHARTS

Growth charts, also known as reference centile charts, were first conceived by Quetelet in the 19th century, and are commonly used to screen the measurements from an individual subject in the context of population values; to this end they are used by medical practitioners, and others, to monitor people's growth. A typical growth chart consists of a family of smooth curves representing a few selected quantiles of the distribution of some physical measurements—height, weight, head circumference, etc.—of the reference population as a function of age. Extreme measurements on the growth chart suggest that the subject should be studied further, to confirm or to rule out an unusual underlying physical condition or disease. The conventional method of constructing growth charts is to get the empirical quantiles of the measurements at a series of time points, and to then fit a smooth polynomial curve using the empirical quantiles—see Hamill et al. (1979). In recent years, a number of

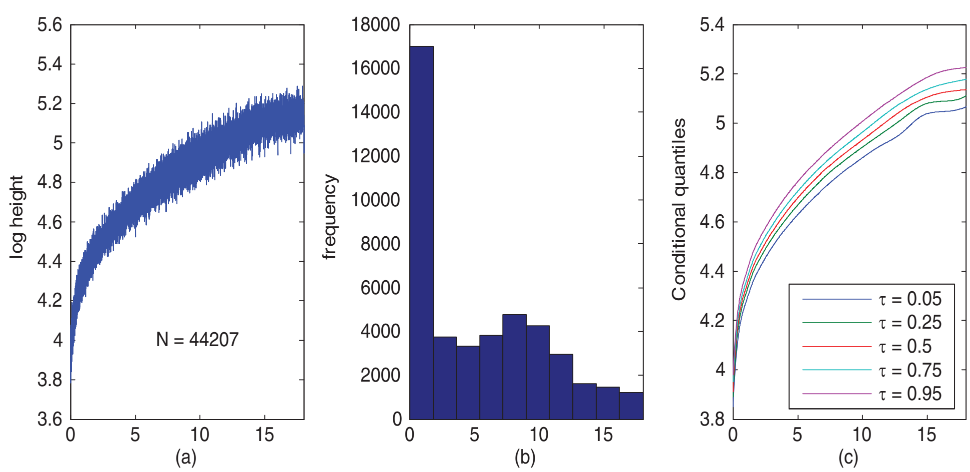

Figure 4. (a) log (height) versus age in full dataset; (b) frequencies of ages; (c) conditional quantile curves computed from full dataset.

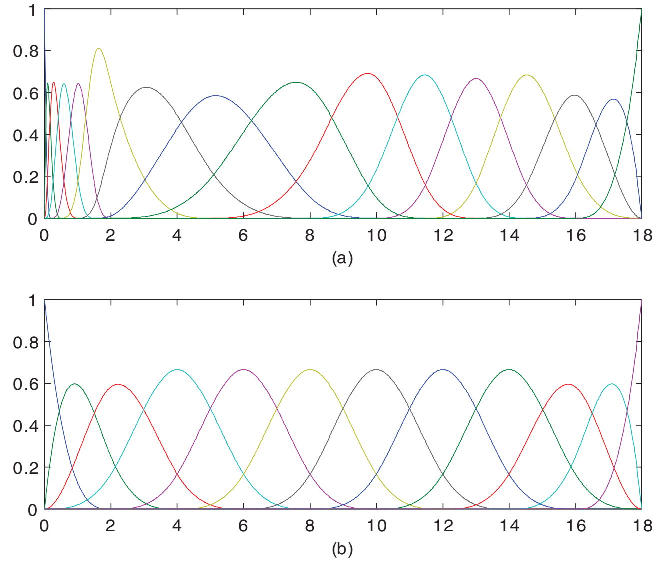

Figure 5. Cubic splines for growth study. (a) "Full" spline basis is of dimension 16; (b) "Reduced" basis of dimension 12 has fewer and different internal knots.

different methods have been developed in the medical statistics literature—see Wei and He (2006) for a review.

A recent method proposed by Wei et al. (2006) is to estimate a family of conditional quantile functions by solving nonparametric quantile regression. In particular, suppose that we want to construct the growth charts for height. As is common practice in pediatrics, we will take the logarithm of height (Y, in centimeters) as response, and age (x, in years), as the covariate. We consider the nonparametric location/scale model

$$Y = \mu(x) + \sigma(x)\varepsilon,$$

where the location function $\mu(x)$ and scale function $\sigma(x)$ satisfy certain smoothness conditions. Given data (y_i, x_i), $i = 1, \ldots, n$ the τth quantile curve can be estimated by minimizing $\sum_{i=1}^{n} \rho_\tau(y_i - \mu(x_i))$.

For growth charts, it is convenient to parameterize the conditional quantile functions as linear combinations of a few basis functions. Particularly convenient for this purpose are cubic B-splines. Given a choice of knots for the B-splines, estimation of the growth charts is a straightforward exercise in parametric linear regression.

The data—see Figure 4, and the detailed description in Pere (2000)—were collected retrospectively from health centres and schools in Finland. To construct the conditional quantile curves in Figure 4(c), for ages from birth to 18 years, we used the entire dataset of size 44,207 and the internal knot sequence

$$\{0.2, 0.5, 1.0, 1.5, 2.0, 5.0, 8.0, 10.0, 11.5, 13.0, 14.5, 16.0\}. \quad (20)$$

This sequence was also used by Wei et al. (2006); see also Kong and Mizera (2012). Spacing of the internal knots is dictated by the need for more flexibility during infancy and in the pubertal growth spurt period. Linear combinations of these functions provide a simple and quite flexible model for the entire curve over [0, 18]. Denoting the selected B-splines by $b_j(x)$ $j = 1, \ldots, p = 16$, we obtain the model (1) with $\mu(x) = \boldsymbol{f}'(\boldsymbol{x})\boldsymbol{\theta}$ for $\boldsymbol{f}(\boldsymbol{x}) = \left(b_1(x), \ldots, b_p(x)\right)'$ and $\boldsymbol{\theta} = \left(\theta_1, \ldots, \theta_p\right)'$. However, due to uncertainty in the selection of knots and to other approximations underlying the model, the designer might well seek protection against departures of the form (3). In this study we will explore how to sample from the available ages to robustly estimate the growth charts of heights.

In computing and assessing the designs we supposed that the experimenter would use the internal knot sequence

$$\{2.0, 4.0, 6.0, 8.0, 10.0, 12.0, 14.0, 16.0\}; \quad (21)$$

one measure of design quality is then the accuracy with which the quantile curves in Figure 4(c), using the "true" model defined by (20), are recovered from the, much smaller, designed sample fitted using (21).

The design space consisted of the $N = 1799$ unique values of x in the original dataset; these span the range [0, 18.0] in increments of 0.01 with only two exceptions. We investigated four types of designs; in all cases illustrated here we used $n = 200$. The first design—"saturated"—places equal weight at each of p points, where $p = 12$ is the number of regression parameters to be estimated to fit the reduced cubic spline basis. The literature provides little guidance on the optimal locations of these points, but we have followed Kaishev (1989) who studied D-optimal

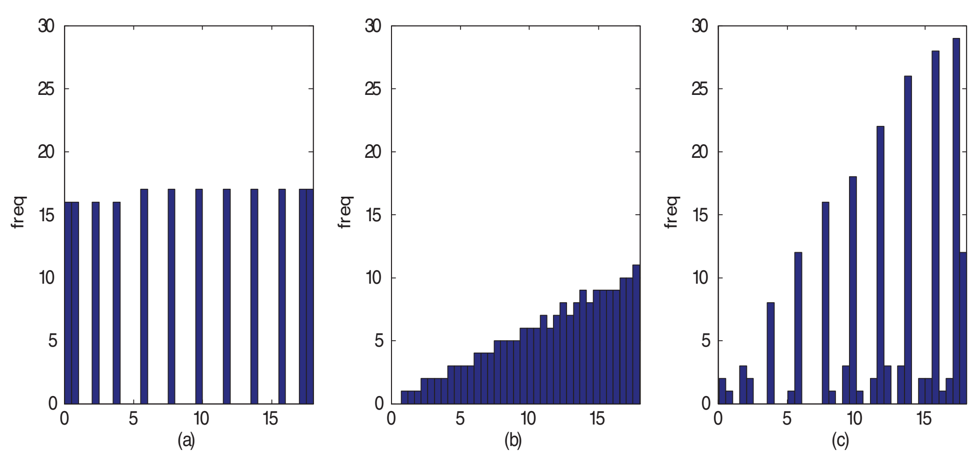

Figure 6. Designs (uniform design not shown) computed for Example 6.1: (a) Saturated design on 12 points; (b) Minbias design; (c) Minimax design with $\nu_0 = 0.5$.

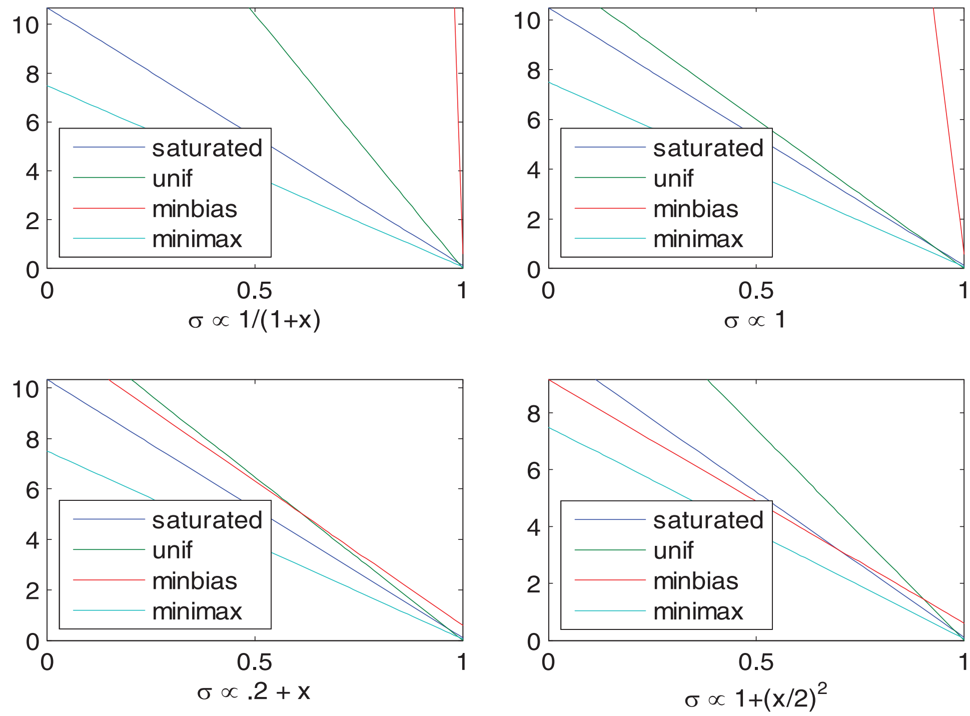

Figure 7. Maximum AMSE $\mathcal{L}_\nu(\xi|\sigma)$ versus ν for various choices of σ and the designs of Example 6.1; the minimax design was tailored to $\nu_0 = 0.5$.

designs for spline models and conjectured that a "near" optimal design places its mass at the p locations at which the individual splines—see Figure 5(b)—attain their maxima. Saturated designs enjoy favored status within optimal design theory, when there is no doubt that the fitted model is in fact the correct one. In this current study they turn out to be quite efficient unless ν is quite large, that is, loss dominated by bias, in which case both the uniform and minimax designs, described below, result in predictions with substantially less bias. As well, the saturated designs are rather poor at recovering the quantile curves from the data gathered at this small number of locations.

The second design is the uniform, implemented as at (9). This has been seen to have minimax properties when the maximum is taken over very broad classes of departures from the nominal model. The third—"minbias"—is as described in Section 4, with designs weights proportional to $\sigma(x)$, again implemented

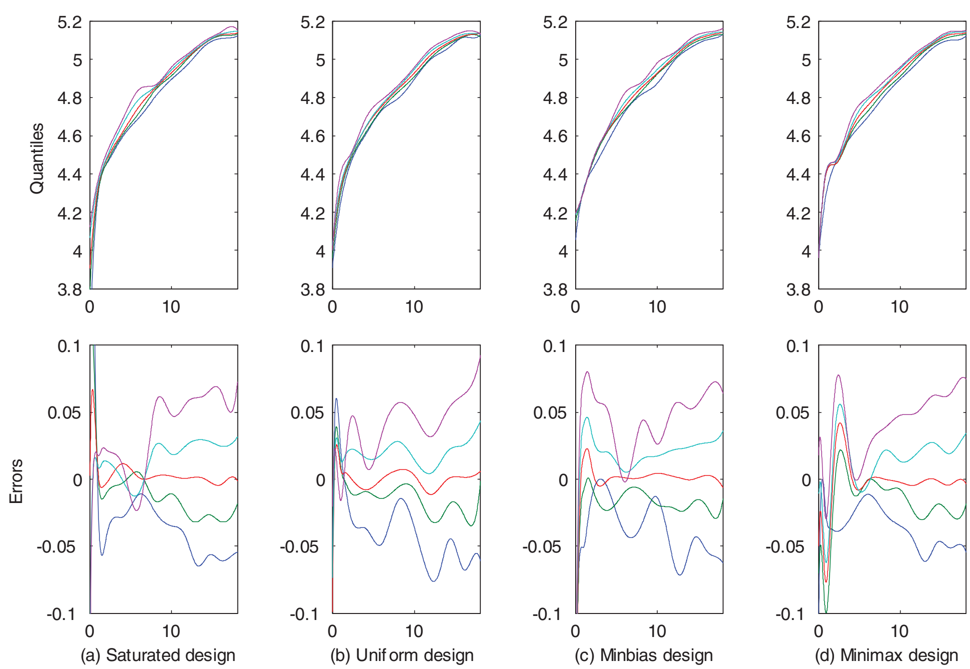

Figure 8. Quantile curves computed in Example 6.1 from the four designs (a)–(d) and reduced spline basis on knots (21), and deviations from those computed using the full dataset and knots (20).

Table 1. Root-mean squared errors (standard errors) for the designs in Example 6.1

	Design			
	Saturated	Uniform	Minbias	Minimax
$\tau = 0.05$	0.061 (0.0009)	0.044 (0.0003)	0.060 (0.0016)	0.052 (0.0008)
$\tau = 0.25$	0.028 (0.0005)	0.020 (0.0002)	0.035 (0.0014)	0.033 (0.0012)
$\tau = 0.50$	0.012 (0.0004)	0.009 (0.0002)	0.026 (0.0015)	0.026 (0.0014)
$\tau = 0.75$	0.022 (0.0002)	0.021 (0.0002)	0.031 (0.0011)	0.038 (0.0015)
$\tau = 0.95$	0.053 (0.0007)	0.046 (0.0003)	0.054 (0.0009)	0.054 (0.0010)

as at (9). It is not possible to implement such a design very accurately when $n < N$, and it will be seen that because of this its minimum bias property is lost. In some cases it does however have attractive behavior with respect to the variance component of the MSE.

The final design—"minimax"—minimizes $\mathcal{L}_\nu(\xi|\sigma)$ at (14) for a particular variance function $\sigma^2(x)$ chosen from those itemized in the captions of Figures 1 and 2. The minimax designs were obtained using a genetic algorithm similar to that described in Welsh and Wiens (2013). The algorithm begins by generating a "population" of 40 designs—the three designs described above and 37 which are randomly generated. Each is assigned a "fitness" value, with the designs having the smallest MSE being the "most fit," and a probabilistic mechanism is introduced by which the most fit members become most likely to be chosen to have "children." The children are formed from the parents in a particular way; with a certain probability they are then subjected to random mutations. In this way the possible parents in each generation are replaced by their children, thus forming the next generation of designs. A feature of the algorithm is that a certain proportion of the members—the most fit 10%—always survive intact; in essence they become their own children. This ensures that the best member of each generation has MSE no larger than that in the previous generation. In all cases we terminated after 1000 generations without improvement.

Example 6.1. We computed the four designs, using the variance function with $\sigma_0(x) \propto 0.2 + x$ and, in the case of the minimax design, a proportion $\nu_0 = 0.5$ of the emphasis placed on bias reduction. See Figure 6. The performance of all designs against all four of the variance functions is illustrated in Figure 7, where the maximum MSE $\mathcal{L}_\nu(\xi|\sigma)$ at (14) is plotted against ν. The efficiency of the minimax design relative to the best of the other three, which we define in terms of the ratio of the corresponding values of $\mathcal{L}_{\nu_0}(\xi|\sigma_0)$, was 1.40—a substantial gain. We then fit quantile curves, for $\tau = 0.05, 0.25, 0.5, 0.75, 0.95$, to the full dataset (Figure 4(c)) and after each design. See Figure 8. For each combination of design and τ, root-mse values were computed as rmse $= \sqrt{\text{mean}(\hat{Y}_{\text{design}} - \hat{Y}_{\text{full}})^2}$, where $\hat{Y}_{\text{full}}$ and $\hat{Y}_{\text{design}}$ refer to predicted values using the full dataset or those obtained from the designs. This required simulating data, which we did as follows. To get data at design point x we sampled from a Normal distribution, with mean given by the value, at x, of the "$\tau = 0.5$" curve in Figure 4(c) and variance $\sigma_Y^2(x)$ estimated from the Y-values, at x, in the original data. This process was carried out 100 times; the rmse values given in Table 1 are the averages of those so obtained, followed by the standard errors in parentheses. The growth and error curves are based on one representative sample. The uniform and minimax designs yielded samples from which the quantile curves were recovered quite accurately; the saturated and minbias designs were generally less successful. In examples not reported here we found however that for substantially larger values of n—for instance $n = 1000$—the minbias design performed as well as the others in this regard.

Example 6.2. We next took $\nu_0 = 0$—all emphasis on variance reduction—but otherwise retained the features of Example 6.1. The saturated, uniform and minbias designs, whose construction does not depend on ν, were thus as in Example 6.1; the minimax design is in Figure 9(a) and again enjoyed a relative efficiency of 1.40 over the best of the others. The plots of the quantile curves—not shown—tell much the same story as those for Example 6.1.

Example 6.3. We then took $\nu_0 = 1$—all emphasis on bias reduction—and obtained the minimax design in Figure 9(b), with a relative efficiency of 1.62. See Table 3, where we give

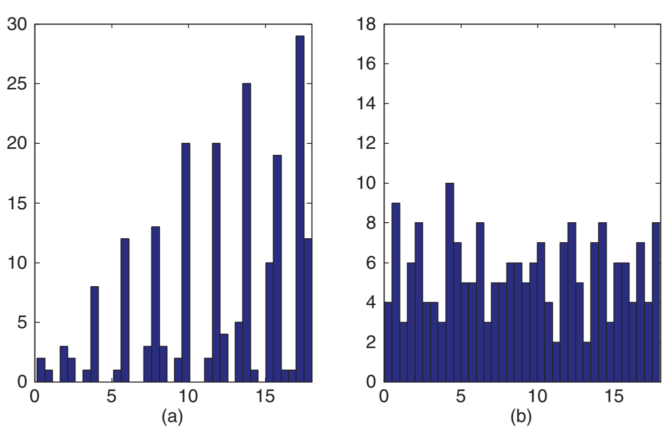

Figure 9. Minimax designs in Examples 6.2 and 6.3; (a) $\nu_0 = 0$; (b) $\nu_0 = 1$.

Table 2. Maximum MSE $\mathcal{L}_\nu(\xi|\sigma_0 \propto .2 + x)$ of the designs in Examples 6.1–6.3

	Design					
				Minimax for:		
	Saturated	Uniform	Minbias	$\nu_0 = 0$	$\nu_0 = 0.5$	$\nu_0 = 1$
$\nu = 0$	10.33	12.94	12.02	7.37	7.40	12.20
$\nu = 0.5$	5.22	6.47	6.31	3.72	3.72	6.10
$\nu = 1$	0.111	0.008	0.600	0.068	0.037	0.005

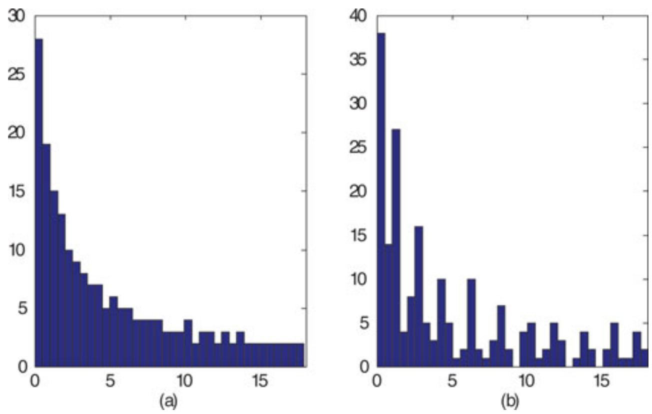

Figure 10. (a) Minbias design and (b) minimax ($\nu_0 = 0.5$) design for Example 6.4; both for $\sigma_0(x) \propto 1/(1+x)$.

the values of $\mathcal{L}_\nu(\xi|\sigma_0)$ for all six designs discussed in Examples 5–7, at $\nu = 0, 0.5, 1$.

Example 6.4. As a final example we reran Example 6.1, but using $\sigma_0(x) \propto 1/(1+x)$. The minimax design had a relative efficiency of 1.17 against the best—the minbias design—of the other three; the efficiency was much greater against the uniform and saturated designs. See Table 2.

7. SUMMARY AND CONCLUDING REMARKS

Dette and Trampisch (2012) studied locally optimal quantile regression designs for nonlinear models, and concluded with a call for future research into the robustness of designs with respect to the model assumptions. In this article we have detailed such research, with specific attention to linear models but with an outline of the modest changes required to address nonlinear models.

Although a number of our methods described here are analytically and numerically complex, some general guidance is possible. One recurring theme of this article is that *uniform* designs are often minimax in sufficiently large classes of the types of departures we consider. It has long been recognized in problems of design for least squares regression that the uniform design plays much the same role as does the median in robust estimation—highly robust if not terribly efficient—and our findings seem to extend this role to quantile regression.

In seeking protection against bias alone, resulting from model misspecification and a particular variance function, designs with weights proportional to the root of the variance function turn out to be minimax against response misspecifications.

Table 3. Maximum MSE $\mathcal{L}_\nu(\xi|\sigma_0 \propto 1/(1+x))$ of the designs in Example 6.4; minimax design uses $\nu_0 = 0.5$

	Design			
	Saturated	Uniform	Minbias	Minimax
$\nu = 0$	10.69	20.81	7.88	6.29
$\nu = 0.5$	5.40	10.41	3.95	3.15
$\nu = 1$	0.111	0.006	0.017	0.016

Uniform designs and minimum bias designs are easily implemented. The more complex design strategies illustrated in Section 3 are more laborious, but it has been seen that a rough description of the outcomes, when there are already available nonrobust designs which minimize the loss at the experimenter's assumed model, is that the robust designs can at least be approximated by taking the replicates prescribed by the nonrobust strategies, and spreading these out into clusters of distinct but nearby design points.

The robust designs obtained here all yield substantial gains in efficiency, as measured in terms of maximum loss, when compared to their competitors—enough to warrant some computational complexity in their construction. As is seen from the plots of the designs—Figures 1, 2, 6, 9, and 10 in particular—a gain in efficiency should be realizable, without a great deal of computation, by merely following the preceding heuristic of clustering replicates, and combining this with design weights suggested by the minimum bias paradigm.

APPENDIX: DERIVATIONS

Mathematical developments for Section 3.1. With definitions $\zeta_i = \xi_i/\sigma_i$, $\boldsymbol{\zeta} = (\zeta_1, \ldots, \zeta_N)'$ and

$$\gamma_0 = \frac{1}{N}\sum_{i=1}^N x_i^2,\ \kappa_1 = \sum_{i=1}^N \zeta_i,\ \omega_1 = \sum_{i=1}^N \zeta_i^2,\ \gamma_1 = \sum_{i=1}^N x_i^2\sigma_i\zeta_i,$$
$$\kappa_2 = \sum_{i=1}^N x_i^2\zeta_i,\ \omega_2 = \sum_{i=1}^N x_i^2\zeta_i^2,$$

(14) becomes $\mathcal{L}_\nu(\xi) = (1-\nu)\{\frac{1}{\kappa_1^2} + \frac{\gamma_0\gamma_1}{\kappa_2^2}\} + \nu\max\{\frac{\omega_1}{\kappa_1^2}, \frac{\gamma_0\omega_2}{\kappa_2^2}\}$. We shall restrict to variance functions for which we can verify that, evaluated at $\{\xi_{*,i}\}_{i=1}^N$,

$$\frac{\omega_1}{\kappa_1^2} \geq \frac{\gamma_0\omega_2}{\kappa_2^2}. \tag{A.1}$$

We thus minimize $(1-\nu)\{\frac{1}{\kappa_1^2} + \frac{\gamma_0\gamma_1}{\kappa_2^2}\} + \nu\frac{\omega_1}{\kappa_1^2}$, first with γ_1, κ_1 and κ_2 fixed; we then minimize over these values. For this we minimize ω_1, subject to

$$\text{(i)} \sum_{i=1}^N x_i^2\sigma_i\zeta_i = \gamma_1,\ \text{(ii)} \sum_{i=1}^N \zeta_i = \kappa_1,\ \text{(iii)} \sum_{i=1}^N x_i^2\zeta_i = \kappa_2,$$
$$\text{(iv)} \sum_{i=1}^N \sigma_i\zeta_i = 1. \tag{A.2}$$

It is sufficient that $\boldsymbol{\zeta} \succeq \mathbf{0}$ (i.e., all elements nonnegative) minimize the convex function

$$\Phi(\boldsymbol{\zeta}, \boldsymbol{\lambda}) = \sum_{i=1}^N [\zeta_i^2 - 2a\{(1+\lambda_1 x_i^2) + \sigma_i(\lambda_2 + \lambda_3 x_i^2)\}\zeta_i],$$

with the multipliers $a(1, \lambda_1, \lambda_2, \lambda_3)'$, prearranged in this convenient manner, chosen to satisfy the side conditions. Since Φ is a sum of univariate, convex functions it is minimized over $\boldsymbol{\zeta} \succeq \mathbf{0}$ at the pointwise positive part $\boldsymbol{\zeta}_0^+ \stackrel{\text{def}}{=} (\zeta_{01}^+, \ldots, \zeta_{0N}^+)'$, where $\boldsymbol{\zeta}_0$ is the stationary point of Φ and $\zeta_{0i}^+ = \max(\zeta_{0i}, 0)$. The calculations yield

$$\zeta_{*i} = \zeta_{*i}(\boldsymbol{\lambda}) = \frac{\{(1+\lambda_1 x_i^2) + \sigma_i(\lambda_2 + \lambda_3 x_i^2)\}^+}{\sum_{i=1}^N \sigma_i\{(1+\lambda_1 x_i^2) + \sigma_i(\lambda_2 + \lambda_3 x_i^2)\}^+}, \tag{A.3}$$

with $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \lambda_3)'$ determined from (i), (ii), and (iii) of (A.2).

We may now minimize over $\boldsymbol{\lambda}$ rather than over $(\gamma_1, \kappa_1, \kappa_2)$, so that the numerical problem is to minimize

$$L(\boldsymbol{\lambda}) = (1-\nu)\left\{\frac{1}{\kappa_1^2} + \frac{\gamma_0\gamma_1}{\kappa_2^2}\right\} + \frac{\nu}{\kappa_1^2}\sum_{i=1}^{N}\zeta_i^2(\boldsymbol{\lambda}),$$

with $\zeta_i(\boldsymbol{\lambda})$ defined by (A.3) and $\gamma_1 = \gamma_1(\boldsymbol{\lambda})$, $\kappa_1 = \kappa_1(\boldsymbol{\lambda})$, $\kappa_2 = \kappa_2(\boldsymbol{\lambda})$ defined by (i), (ii), and (iii) of (A.2). After doing this with a numerical constrained minimizer we check (A.1). Then $\xi_{*i} = \sigma_i\zeta_{*i}$.

In Figure 1 we have illustrated only some representative variance functions for which (A.1) holds. When it does not, one can minimize instead $(1-\nu)\{\frac{1}{\kappa_1^2} + \frac{\gamma_0\gamma_1}{\kappa_2^2}\} + \nu\frac{\gamma_0\omega_2}{\kappa_2^2}$ and then check that, at the optimal design, $\frac{\gamma_0\omega_2}{\kappa_2^2} \geq \frac{\omega_1}{\kappa_1^2}$. If this also fails, then a more complex method which is however guaranteed to succeed is that of Daemi and Wiens (2013), used in Section 3.2. □

Proof of Theorem 2. By (12) we are to find

$$\max_{\Delta_0}\text{AMSE} = \frac{\tau(1-\tau)}{g_\varepsilon^2(0)}\text{tr}\left(\boldsymbol{A}\boldsymbol{P}_1^{-1}\boldsymbol{P}_0\boldsymbol{P}_1^{-1}\right) + \max_{\Delta_0}\left[\boldsymbol{\mu}_0'\boldsymbol{P}_1^{-1}\boldsymbol{A}\boldsymbol{P}_1^{-1}\boldsymbol{\mu}_0 + N^{-1}\sum_{i=1}^{N}\delta_0^2(\boldsymbol{x}_i)\right].$$

We use methods introduced in Fang and Wiens (2000). We first represent the design by a diagonal matrix $\boldsymbol{D}_\xi$ with diagonal elements $\{\xi_i\}$. Define $\boldsymbol{D}_\sigma$ to be the diagonal matrix with diagonal elements $\{\sigma(\boldsymbol{x}_i)\}$. Let $\boldsymbol{Q}_1$ be an $N \times p$ matrix whose columns form an orthogonal basis for the column space of the matrix $\boldsymbol{F}$ with rows $\{\boldsymbol{f}'(\boldsymbol{x}) \mid \boldsymbol{x}\in\mathcal{X}\}$—recall that this is "Q" in the QR-decomposition of $\boldsymbol{F}$. Then $\boldsymbol{F} = \boldsymbol{Q}_1\boldsymbol{R}$ for a $p \times p$, nonsingular triangular matrix $\boldsymbol{R}$. 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 $[\boldsymbol{Q}_1 \vdots \boldsymbol{Q}_2]$ is an orthogonal matrix and $\boldsymbol{\delta}_0 = (\delta_0(\boldsymbol{x}_1), \ldots, \delta_0(\boldsymbol{x}_N))'$ is, by (i) of (8a), of the form $\boldsymbol{\delta}_0 = \eta\boldsymbol{Q}_2\boldsymbol{c}$, where $\|\boldsymbol{c}\| \leq 1$. In these terms $\boldsymbol{A} = N^{-1}\boldsymbol{R}'\boldsymbol{R}$ and from (10a)–(10c),

$$\boldsymbol{\mu}_0 = \eta\boldsymbol{R}'\boldsymbol{Q}_1'\boldsymbol{D}_\sigma^{-1}\boldsymbol{D}_\xi\boldsymbol{Q}_2\boldsymbol{c},\quad \boldsymbol{P}_0 = \boldsymbol{R}'\boldsymbol{Q}_1'\boldsymbol{D}_\xi\boldsymbol{Q}_1\boldsymbol{R},$$
$$\boldsymbol{P}_1 = \boldsymbol{R}'\boldsymbol{Q}_1'\boldsymbol{D}_\sigma^{-1}\boldsymbol{D}_\xi\boldsymbol{Q}_1\boldsymbol{R}.$$

Thus

$$\begin{aligned}\max_{\Delta_0}&\left[\boldsymbol{\mu}_0'\boldsymbol{P}_1^{-1}\boldsymbol{A}\boldsymbol{P}_1^{-1}\boldsymbol{\mu}_0 + N^{-1}\sum_{i=1}^{N}\delta_0^2(\boldsymbol{x}_i)\right]\\
&= \frac{\eta^2}{N}\max_{\|\boldsymbol{c}\|\leq 1}[(\boldsymbol{c}'\boldsymbol{Q}_2'\boldsymbol{D}_\xi\boldsymbol{D}_\sigma^{-1}\boldsymbol{Q}_1\boldsymbol{R})(\boldsymbol{P}_1^{-1}\boldsymbol{R}'\boldsymbol{R}\boldsymbol{P}_1^{-1})\\
&\quad\times(\boldsymbol{R}'\boldsymbol{Q}_1'\boldsymbol{D}_\sigma^{-1}\boldsymbol{D}_\xi\boldsymbol{Q}_2\boldsymbol{c}) + \boldsymbol{c}'\boldsymbol{Q}_2'\boldsymbol{Q}_2\boldsymbol{c}]\\
&= \frac{\eta^2}{N}ch_{\max}[\boldsymbol{Q}_2'\{\boldsymbol{D}_\xi\boldsymbol{D}_\sigma^{-1}\boldsymbol{Q}_1\boldsymbol{R}\boldsymbol{P}_1^{-1}\boldsymbol{R}'\boldsymbol{R}\boldsymbol{P}_1^{-1}\boldsymbol{R}'\boldsymbol{Q}_1'\boldsymbol{D}_\sigma^{-1}\boldsymbol{D}_\xi\\
&+\mathbf{I}_{N-p}\}\boldsymbol{Q}_2].\end{aligned}$$

Some algebra, followed by a return to the original parameterization, results in (14). □

Proof of Theorem 3. This parallels the proof of Theorem 1 of Wiens (1992), and can also be obtained by taking limits, as $N \to \infty$, in Theorem 2. □

Derivation of (15). The, rather lengthy, calculations for this section are available in Kong and Wiens (2014). As in Section 3.1, we consider symmetric designs and variance functions: $m(x) = m(-x)$ and $\sigma(x) = \sigma(-x)$. In terms of

$$\mu_i = \int_{-1}^{1} x^i m(x)dx,\ \kappa_i = \int_{-1}^{1} x^i\frac{m(x)}{\sigma(x)}dx,\ \omega_i = \int_{-1}^{1} x^i\left(\frac{m(x)}{\sigma(x)}\right)^2 dx,$$

we define $\pi = 2/\left(\kappa_4\kappa_0 - \kappa_2^2\right)^2$, $\phi_{002} = \pi/\left(3\kappa_2^2\right)$ and

$$\phi_{110} = \pi\left[\kappa_4^2 - \frac{1}{3}\kappa_4\kappa_2\right], \phi_{112} = \pi\left[\frac{1}{3}\left(\kappa_4\kappa_0 + \kappa_2^2\right) - 2\kappa_4\kappa_2\right],$$
$$\phi_{114} = \pi\left[\kappa_2^2 - \frac{1}{3}\kappa_2\kappa_0\right],$$
$$\phi_{120} = \pi\left[\frac{1}{3}\kappa_2^2 - \kappa_4\kappa_2\right], \phi_{122} = \pi\left[\kappa_4\kappa_0 + \kappa_2^2 - \frac{2}{3}\kappa_2\kappa_0\right],$$
$$\phi_{124} = \pi\left[\frac{1}{3}\kappa_0^2 - \kappa_2\kappa_0\right],$$
$$\phi_{210} = \pi\left[\frac{1}{3}\kappa_4^2 - \frac{1}{5}\kappa_4\kappa_2\right], \phi_{212} = \pi\left[\frac{1}{5}\left(\kappa_4\kappa_0 + \kappa_2^2\right) - \frac{2}{3}\kappa_4\kappa_2\right],$$
$$\phi_{214} = \pi\left[\frac{1}{3}\kappa_2^2 - \frac{1}{5}\kappa_2\kappa_0\right],$$
$$\phi_{220} = \pi\left[\frac{1}{5}\kappa_2^2 - \frac{1}{3}\kappa_4\kappa_2\right], \phi_{222} = \pi\left[\frac{1}{3}\left(\kappa_4\kappa_0 + \kappa_2^2\right) - \frac{2}{5}\kappa_2\kappa_0\right],$$
$$\phi_{224} = \pi\left[\frac{1}{5}\kappa_0^2 - \frac{1}{3}\kappa_2\kappa_0\right].$$

We then calculate that

$$\text{tr}(\mathbf{A}\boldsymbol{T}_0) \stackrel{\text{def}}{=} \rho_0(m) = [\phi_{110} + \phi_{220}] + [\phi_{002} + \phi_{112} + \phi_{222}]\mu_2 + [\phi_{114} + \phi_{224}]\mu_4,$$

and that

$$\mathbf{A}\boldsymbol{T}_2 = \begin{pmatrix}\phi_{110}\omega_0 + \phi_{112}\omega_2 + \phi_{114}\omega_4 & 0 & \phi_{120}\omega_0 + \phi_{122}\omega_2 + \phi_{124}\omega_4\\ 0 & \phi_{002}\omega_2 & 0\\ \phi_{210}\omega_0 + \phi_{212}\omega_2 + \phi_{214}\omega_4 & 0 & \phi_{220}\omega_0 + \phi_{222}\omega_2 + \phi_{224}\omega_4\end{pmatrix},$$

whose characteristic roots are $\rho_1(m) = \phi_{002}\omega_2$ and the two roots of

$$\begin{pmatrix}\phi_{110}\omega_0 + \phi_{112}\omega_2 + \phi_{114}\omega_4 & \phi_{120}\omega_0 + \phi_{122}\omega_2 + \phi_{124}\omega_4\\ \phi_{210}\omega_0 + \phi_{212}\omega_2 + \phi_{214}\omega_4 & \phi_{220}\omega_0 + \phi_{222}\omega_2 + \phi_{224}\omega_4\end{pmatrix} \stackrel{\text{def}}{=} \begin{pmatrix}\psi_{11} & \psi_{12}\\ \psi_{21} & \psi_{22}\end{pmatrix}.$$

Of these two roots, one is uniformly greater than the other, and is

$$\rho_2(m) = \frac{\psi_{11} + \psi_{22}}{2} + \left\{\left(\frac{\psi_{11} - \psi_{22}}{2}\right)^2 + \psi_{12}\psi_{21}\right\}^{1/2}.$$

Thus the loss is $\max(\mathcal{L}_1(m), \mathcal{L}_2(m))$, where $\mathcal{L}_k(m) = (1-\nu)\rho_0(m) + \nu\rho_k(m),\ k = 1, 2$.

We apply Theorem 1 of Daemi and Wiens (2013), by which we may proceed as follows. We first find a density m_1 minimizing $\mathcal{L}_1(m)$ in the class of densities for which $\mathcal{L}_1(m) = \max(\mathcal{L}_1(m), \mathcal{L}_2(m))$, and a density m_2 minimizing $\mathcal{L}_2(m)$ in the class of densities for which $\mathcal{L}_2(m) = \max(\mathcal{L}_1(m), \mathcal{L}_2(m))$. Then the optimal design ξ_* has density

$$m_* = \begin{cases} m_1, & \text{if } \mathcal{L}_1(m_1) \leq \mathcal{L}_2(m_2),\\ m_2, & \text{if } \mathcal{L}_2(m_2) \leq \mathcal{L}_1(m_1).\end{cases}$$

The two minimizations are first carried out with $\mu_2, \mu_4, \kappa_0, \kappa_2, \kappa_4$ held fixed, thus fixing all ϕ_{ijk} and $\rho_0(m)$. Under these constraints $\mathcal{L}_1(m_1) \leq \mathcal{L}_2(m_2)$ iff $\rho_1(m_1) \leq \rho_2(m_2)$.

With the aid of Lagrange multipliers we find that both m_1 and m_2 are of the form (15). The ten constants a_{ij} forming $\mathbf{a}$ are chosen to minimize the loss subject to the side conditions, but it is now numerically simpler to minimize $\mathcal{L}_\nu(\xi|\sigma)$ at (14) directly over $\mathbf{a}$, subject to $\int_{-1}^{1} m(x;\mathbf{a})dx = 1$.

The density $m(x;\mathbf{a})$ is overparameterized, and when $\sigma(\cdot)$ is nonconstant we take $a_{01} = 1$. In the homogeneous case we take $a_{02} = 1$ and also $a_{i1} \equiv 0$ and $a_{00} = 0$. □

Proof of Proposition 1. We give the proof of (i); that of (ii) is similar. For $i = 1, \ldots, N$ define $\boldsymbol{b}(\boldsymbol{x}_i) = (\mathbf{M}_p^{-1}p(\boldsymbol{x}_i) - \mathbf{M}_1^{-1})\boldsymbol{f}(\boldsymbol{x}_i)I(\boldsymbol{x}_i \in \chi_0)$.

Then

$$0 \preceq \sum_{i=1}^{N} \boldsymbol{b}(\boldsymbol{x}_i)\boldsymbol{b}'(\boldsymbol{x}_i) = \mathbf{M}_p^{-1}\mathbf{M}_{p^2}\mathbf{M}_p^{-1} - \mathbf{M}_1^{-1}.$$

Proof of Lemma 1. Write

$$\mathcal{L}_\nu(\xi|r) = (1-\nu)N\frac{\operatorname{tr}\left(\boldsymbol{A}\boldsymbol{S}_1^{-1}(r)\,\boldsymbol{S}_0\boldsymbol{S}_1^{-1}(r)\right)}{\sum_{\xi_i>0}\xi_i^r} + \nu ch_{\max}\left(\boldsymbol{A}\boldsymbol{S}_1^{-1}(r)\,\boldsymbol{S}_2(r)\,\boldsymbol{S}_1^{-1}(r)\right),$$

and note that $\mathcal{L}_\nu(\xi_k|r) = (1-\nu)N\operatorname{tr}(\boldsymbol{A}\boldsymbol{A}_\xi^{-1}) + \nu ch_{\max}(\boldsymbol{A}\boldsymbol{A}_\xi^{-1})$, independently of r. Thus it suffices to show that for some $r = r_\xi$,

$$\mathcal{L}_\nu\left(\xi|r_\xi\right) \geq (1-\nu)\,N\operatorname{tr}\left(\boldsymbol{A}\boldsymbol{A}_\xi^{-1}\right) + \nu ch_{\max}\left(\boldsymbol{A}\boldsymbol{A}_\xi^{-1}\right). \tag{A.5}$$

In fact $r_\xi = 1$ serves the purpose. To see this note that by Proposition 1,

$$\frac{\operatorname{tr}(\boldsymbol{A}\boldsymbol{S}_1^{-1}(1)\boldsymbol{S}_0\boldsymbol{S}_1^{-1}(1))}{\sum_{\xi_i>0}\xi_i} \geq \operatorname{tr}(\boldsymbol{A}\boldsymbol{A}_\xi^{-1}),$$

and that for *any* r, $\boldsymbol{S}_1^{-1}(r)\boldsymbol{S}_2(r)\boldsymbol{S}_1^{-1}(r) \succeq \boldsymbol{A}_\xi^{-1}$, so that also $ch_{\max}(\boldsymbol{A}\boldsymbol{S}_1^{-1}(r)\boldsymbol{S}_2(r)\boldsymbol{S}_1^{-1}(r)) \geq ch_{\max}(\boldsymbol{A}\boldsymbol{A}_\xi^{-1})$. This establishes (A.4) with $r_\xi = 1$. □

[Received July 2014. Revised September 2014.]

REFERENCES

Behl, P., Claeskes, G., and Dette, H. (2014), "Focussed Model Selection in Quantile Regression," *Statistica Sinica*, 24, 601–624. [233]

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. [237]

Bischoff, W. (2010), "An Improvement in the Lack-of-Fit Optimality of the (Absolutely) Continuous Uniform Design in Respect of Exact Designs," in *Proceedings of the 9th International Workshop in Model-Oriented Design and Analysis (moda9)*, eds. Giovagnoli, Atkinson, and Torsney, Berlin Heidelberg: Springer-Verlag. [237]

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. [233,238]

Cook, R. D., and Wong, W. K. (1994), "On the Equivalence of Constrained and Compound Optimal Designs," *Journal of the American Statistical Association*, 89, 687–692. [239]

Daemi, M., and Wiens, D. P. (2013), "Techniques for the Construction of Robust Regression Designs," *The Canadian Journal of Statistics*, 41, 679–695. [238,244]

Dette, H., and Trampisch, M. (2012), "Optimal Designs for Quantile Regression Methods," *Journal of the American Statistical Association*, 107, 1140–1151. [233,235,243]

Fang, K. T., and Wang, Y. (1994), *Number-Theoretical Methods in Statistics*, London, UK: Chapman and Hall. [235]

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. [236,244]

Hamill, P. V. V., Dridzd, T. A., Johnson, C. L., Reed, R. B., Roche, A. F., and Moore, W. M. (1979), "Physical Growth: National Center for Health Statistics percentiles," *American Journal of Clinical Nutrition*, 32, 607–629. [239]

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. [237]

Huber, P. J. (1964), "Robust Estimation of a Location Parameter," *The Annals of Mathematical Statistics*, 35, 73–101. [233]

——— (1975), "Robustness and Designs," in *A Survey of Statistical Design and Linear Models*, ed. J. N. Srivastava, Amsterdam: North Holland, pp. 287–303. [233]

——— (1981), *Robust Statistics*, New York: Wiley. [233]

Kaishev, V. K. (1989), "Optimal Experimental Designs for the B-Spline Regression," *Computational Statistics & Data Analysis*, 8, 39–47. [240]

Knight, K. (1998), "Limiting Distributions for l_1 Estimators Under General Conditions," *The Annals of Statistics*, 26, 755–770. [235]

Koenker, R. (2005), *Quantile Regression*, Cambridge, UK: Cambridge University Press. [235]

Koenker, R., and Bassett, G. (1978), "Regression Quantiles," *Econometrica*, 46, 33–50. [233]

Kong, L., and Mizera, I. (2012), "Quantile Tomography: Using Quantiles With Multivariate Data," *Statistica Sinica*, 22, 1589–1610. [240]

Kong, L., and Wiens, D. P. (2014), "Robust Quantile Regression Designs," University of Alberta Department of Mathematical and Statistical Sciences Technical Report S129, available at *http://www.stat.ualberta.ca/~wiens/homepage/pubs/TR S129.pdf*. [234,235,238,244]

Li, K. C. (1984), "Robust Regression Designs When the Design Space Consists of Finitely Many Points," *The Annals of Statistics*, 12, 269–282. [233]

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. [233]

Ma, Y., and Wei, Y. (2012), "Analysis on Censored Quantile Residual Life Model via Spline Smoothing," *Statistica Sinica*, 22, 47–68. [233]

Maronna, R. A., and Yohai, V. J. (1981), "Asymptotic Behaviour of General M-Estimates for Regression and Scale With Random Carriers," *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 58, 7–20. [233]

Martínez-Silva, I., Roca-Pardiñas, J., Lustres-Pérez, V., Lorenzo-Arribas, A., and Cadaro-Suárez, C. (2013), "Flexible Quantile Regression Models: Application to the Study of the Purple Sea Urchin," *SORT—Statistics and Operations Research Transactions*, 37, 81–94. [233]

Pere, A. (2000), "Comparison of Two Methods of Transforming Height and Weight to Normality," *Annals of Human Biology*, 27, 35–45. [240]

Rubia, A., and Sanchis-Marco, L. (2013), "On Downside Risk Predictability Through Liquidity and Trading Activity: A Dynamic Quantile Approach," *International Journal of Forecasting*, 29, 202–219. [233]

Shi, P., Ye, J., and Zhou, J. (2003), "Minimax Robust Designs for Misspecified Regression Models," *The Canadian Journal of Statistics*, 31, 397–414. [237]

Simpson, D. G., Ruppert, D., and Carroll, R. J. (1992), "On One-Step GM Estimates and Stability of Inferences in Linear Regression," *Journal of the American Statistical Association*, 87, 439–450. [233]

Wei, Y., and He, X. (2006), "Discussion Article: Conditional Growth Charts," *The Annals of Statistics*, 34, 2069–2097. [240]

Wei, Y., Pere, A., Koenker, R., and He, X. (2006), "Quantile Regression Methods for Reference Growth Charts," *Statistics in Medicine*, 25, 1369–1382. [240]

Welsh, A. H., and Wiens, D. P. (2013), "Robust Model-Based Sampling Designs," *Statistics and Computing*, 23, 689–701. [242]

Wiens, D. P. (1991), "Designs for Approximately Linear Regression: Two Optimality Properties of Uniform Designs," *Statistics and Probability Letters*, 12, 217–221. [237]

——— (1992), "Minimax Designs for Approximately Linear Regression," *Journal of Statistical Planning and Inference*, 31, 353–371. [233,237,244]

——— (2000), "Robust Weights and Designs for Biased Regression Models: Least Squares and Generalized M-Estimation," *Journal of Statistical Planning and Inference*, 83, 395–412. [233]

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. [233]

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, 84–292. [233]

Xu, X., and Yuen, W. K. (2011), "Applications and Implementations of Continuous Robust Designs," *Communications in Statistics—Theory and Methods*, 40, 969–988. [235]