Article

SMMR
STATISTICAL METHODS IN MEDICAL RESEARCH

Statistical Methods in Medical Research
2021, Vol. 30(1) 221–232

DOI: 10.1177/0962280220948159
journals.sagepub.com/home/smm
SAGE

Model-robust designs for nonlinear quantile regression

Selvakkadunko Selvaratnam, Linglong Kong and Douglas P Wiens

Abstract

We construct robust designs for nonlinear quantile regression, in the presence of both a possibly misspecified nonlinear quantile function and heteroscedasticity of an unknown form. The asymptotic mean-squared error of the quantile estimate is evaluated and maximized over a neighbourhood of the fitted quantile regression model. This maximum depends on the scale function and on the design. We entertain two methods to find designs that minimize the maximum loss. The first is local – we minimize for given values of the parameters and the scale function, using a sequential approach, whereby each new design point minimizes the subsequent loss, given the current design. The second is adaptive – at each stage, the maximized loss is evaluated at quantile estimates of the parameters, and a kernel estimate of scale, and then the next design point is obtained as in the sequential method. In the context of a Michaelis–Menten response model for an estrogen/hormone study, and a variety of scale functions, we demonstrate that the adaptive approach performs as well, in large study sizes, as if the parameter values and scale function were known beforehand and the sequential method applied. When the sequential method uses an incorrectly specified scale function, the adaptive method yields an, often substantial, improvement. The performance of the adaptive designs for smaller study sizes is assessed and seen to still be very favourable, especially so since the prior information required to design sequentially is rarely available.

Keywords

Adaptive, kernel, minimax, model robustness, quantile, sequential

1 Introduction and summary

Quantile regression analyses have become common among investigators seeking summaries of their data that are richer and more complete than can be afforded by a standard regression, which returns estimates only of the means, conditional on the covariates. See Koenker[1] and Koenker[2] for an in-depth review. Recent application areas include medical reference charts, survival analysis, financial economics, environmental modelling and the detection of heteroscedasticity – see Yu et al.[3], Geraci[4] and Farcomeni and Geraci[5] for details of these and other novel applications of quantile regression and discussions of some of the computational challenges involved.

The role of a properly constructed design, especially in medical studies, to minimize research costs while maintaining statistical efficiency cannot be overemphasized. By a *design* for an experimental study, we mean the choice of values of the predictor variables, at which a response will be obtained and measured. One hopes that the estimates obtained, upon implementing the design, will be accurate and efficient, i.e. have small biases and small variances. Whether or not these goals are simultaneously attainable depends very much on the accuracy of the model, relating predictors to responses that the experimenter fits. In 'classical' optimal design theory, the

Department of Mathematical and Statistical Sciences, University of Alberta, Alberta, Canada

Corresponding author:
Douglas P Wiens, Department of Mathematical and Statistical Sciences, University of Alberta, Alberta, Canada.
Email: doug.wiens@ualberta.ca

designer claims complete faith in the accuracy of this fitted model and under this assumption can often derive designs yielding unbiased estimates with maximal efficiency.

But it has repeatedly been seen in robustness studies that an 'optimal' procedure for an even slightly incorrect model may be far from optimal – perhaps even worse than useless – under seemingly minor departures from the fitted model. Thus, we instead design for 'model-robustness' – while the experimenters will continue to fit and estimate their chosen model, the designs given here afford protection, in a 'minimax' sense, against the losses arising from misspecifications of this model. To incorporate errors arising both from bias and from natural variation, we measure the loss by the mean squared error (mse) of the predicted values and seek designs that minimize the maximum mse, as the 'true' model ranges over a given neighbourhood of the fitted model. See Wiens[6] for a comprehensive review.

Dette and Trampisch[7] introduced the construction of optimal designs for quantile regression, considering in particular nonlinear response models. For nonlinear responses, the procedures typically depend on the parameters being estimated, and so the designs of Dette and Trampisch[7] were *local* in nature, i.e. tuned to, and optimal for, fixed parameter values, assumed known a priori. Kong and Wiens[8] constructed *robust* quantile regression designs – robust in the manner described above, within certain neighbourhoods of the fitted but possibly misspecified *linear* response function. Both in Dette and Trampisch[7] and Kong and Wiens,[8] the possibly heteroscedastic variances of the observations were assumed known.

In this article, we extend both Dette and Trampisch[7] and Kong and Wiens[8] and we consider a nonlinear response model, heteroscedasticity of an unknown form and a possibly misspecified quantile response function. Our main observation is that appreciable gains can be made by taking an *adaptive* approach in the design construction. For this, we begin by outlining a *sequential* method, in which we design locally, choosing design points one at a time in order to minimize the subsequent loss, conditional on the current design. This is presented for purposes of comparison and to motivate the adaptive approach, in which the updating criterion uses estimates of both regression and scale, computed from the current design.

Adaptive methods for nonlinear models, without regard to robustness, have been developed by Ford and Silvey[9] and Wu[10], among others. For some particular applications, see Santos and Santos[11] and Guest and Curtis.[12] Robustness has been incorporated into the investigations by Sinha and Wiens[13] and Wiens.[14] Adaptive designs can be modified mid-stream, adding an element of flexibility to a study. Details of planning and implementing adaptive clinical trials are discussed in Müller and Schäfer,[15] Konig,[16] Levin et al.[17] and Magirr et al.[18]

In the second section of this article, we discuss nonlinear quantile regression and introduce our robust design requirements. These lead to designs defined as minimizers of a certain loss function that has been maximized over the possible departures from the fitted model. In the third section, we obtain these minimizing designs, using each of the two approaches – sequential and adaptive – described above. We do this in the context of an investigation of an estrogen receptor assay of human uterine cytosol fraction carried out by Cressie and Keightley.[19]

The results are very encouraging. In our simulations, we find that, when the sequential method uses the correct scale function, it is still only slightly better, in large study sizes, than the adaptive method that must estimate scale. When the sequential method uses an incorrect scale function, the adaptive method can yield substantial improvements. The performance of the adaptive designs for smaller study sizes is assessed and seen to still be very favourable, especially so since the prior information required to design sequentially is rarely available.

2 Nonlinear quantile regression and design robustness

We suppose that an experimenter intends to make observations on random variables Y with structure

$$Y = F(\boldsymbol{x}; \boldsymbol{\theta}_\tau) + \sigma(\boldsymbol{x})\varepsilon \tag{1}$$

for a function F of arguments $\boldsymbol{x}_{q\times 1}$ to be chosen by the experimenter from a 'design space' $\chi = \{\boldsymbol{x}_1, \boldsymbol{x}_2 \ldots, \boldsymbol{x}_N\}$, and unknown parameters $\boldsymbol{\theta}_\tau : p \times 1$. For a fixed $\tau \in (0, 1)$, we seek to estimate $\boldsymbol{\theta}_\tau$, which is such that $F(\boldsymbol{x}; \boldsymbol{\theta}_\tau)$ is the conditional τ-quantile of Y, given $\boldsymbol{x}$

$$\tau = G_\varepsilon(0) = G_{Y|\boldsymbol{x}}(F(\boldsymbol{x}; \boldsymbol{\theta}_\tau))$$

We assume that the errors ε are i.i.d., with distribution function G_ε and density g_ε and that the variance function $\sigma^2(\boldsymbol{x})$ is strictly positive on χ. Our asymptotics require that g_ε be Lipschitz continuous and continuously

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

$$Y = F(\boldsymbol{x}; \boldsymbol{\theta}_\tau) + \delta(\boldsymbol{x}) + \sigma(\boldsymbol{x})\varepsilon$$

for some 'small' model error δ. Define $\delta^*(\boldsymbol{x}) = \delta(\boldsymbol{x})/\sigma(\boldsymbol{x})$. We impose the conditions

$$\text{(i) } N^{-1}\sum_{i=1}^{N} \delta^{*^2}(\boldsymbol{x}_i) \leq \eta^2/n, \quad \text{(ii) } N^{-1}\sum_{\text{i}=1}^{N} \sigma^2(\boldsymbol{x}_\text{i}) = 1 \tag{2}$$

for a constant η. The first of these determines the order of the bias, ensuring that it and standard error remain of the same asymptotic order in the mse. Condition (2ii) is without loss of generality, since a multiplicative constant can be absorbed into the scale of G_ε. The condition then fixes this constant, thus ensuring the identifiability of the variance function.

The experimenter, acting as though $\delta \equiv 0$, computes

$$\hat{\boldsymbol{\theta}}_{n\tau} = \arg\min_{\boldsymbol{\theta}} \sum_{i=1}^{n} \rho_\tau\big(y_i - F\big(\boldsymbol{x}_{(i)}; \boldsymbol{\theta}\big)\big) \tag{3}$$

from a sample $\{\boldsymbol{x}_{(i)}, y_i\}_{i=1}^n$. Here, $\rho_\tau(\cdot)$ is the 'check' function $\rho_\tau(r) = r(\tau - I(r < 0))$, with derivative $\psi_\tau(r) = \tau - I(r < 0)$. The target parameter is defined by

$$\boldsymbol{\theta}_\tau = \arg\min_{\boldsymbol{\theta}} \frac{1}{N}\sum_{i=1}^{N} E_{Y|\boldsymbol{x}_i}[\rho_\tau(Y - F(\boldsymbol{x}_i; \boldsymbol{\theta}))] \tag{4}$$

implying

$$\frac{1}{N}\sum_{i=1}^{N} E_{Y|\boldsymbol{x}_i}[\psi_\tau(Y - F(\boldsymbol{x}_i; \boldsymbol{\theta}))]\boldsymbol{f}(\boldsymbol{x}_i, \boldsymbol{\theta}) = \mathbf{0}$$

where we define $\boldsymbol{f}(\boldsymbol{x}, \boldsymbol{\theta}) = \partial F(\boldsymbol{x}; \boldsymbol{\theta})/\partial \boldsymbol{\theta} : p \times 1$. Condition (4) leads to the orthogonality requirement

$$\frac{1}{N}\sum_{i=1}^{N} \delta^*(\boldsymbol{x}_i) f(\boldsymbol{x}_i, \boldsymbol{\theta}_\tau) = \mathbf{0} \tag{5}$$

in the presence of (2ii), this ensures the identifiability of $\delta(\boldsymbol{x})$.

Let $\boldsymbol{\xi}$ be the design measure, with $\xi_i = \boldsymbol{\xi}(\boldsymbol{x}_i)$ giving the fraction of observations made at $\boldsymbol{x}_i$. Then the estimate satisfies

$$\sum_{i=1}^{N} \psi_\tau(Y - F(\boldsymbol{x}_i; \boldsymbol{\theta}))\boldsymbol{f}(\boldsymbol{x}_i, \boldsymbol{\theta})\boldsymbol{\xi}(\boldsymbol{x}_i) = \mathbf{0}$$

The asymptotic normality of $\hat{\boldsymbol{\theta}}_{n\tau}$, conditional on the design measure $\boldsymbol{\xi}$, is given in Theorem 1 below. We briefly summarize some history of such results, all derived with an eye to *estimation* rather than design. Jurečková and Prochazka[20] established the asymptotic normality of quantile regression estimators in *correctly specified*, homoscedastic, *nonlinear* models with i.i.d. errors; Oberhofer and Haupt[21] extended these results under heterogeneity and mixing conditions. Kim and White[22] considered *misspecified*, *linear* quantile regression models. They show that a Fisher-consistent estimate retains the same asymptotic properties as does the quantile regression estimate in correctly specified models.

Komunjer[23] relates consistent estimation in correctly specified models to quasi-likelihood estimation in 'tick-exponential' families and goes on to show that, in misspecified models, the quasi-likelihood estimate is consistent

for a 'pseudo-true' parameter value, maximizing the pseudo-likelihood, around which it is asymptotically normally distributed. In a somewhat similar vein, Angrist et al.[24] show that *linear* quantile regression provides a weighted least squares approximation to a possibly nonlinear conditional quantile function and study some implications of this observation.

In contrast, Theorem 1 treats nonlinear models with possible misspecification of the form outlined above. The bias is presented explicitly, emphasizing its dependence on the *design* and thus setting the stage for design optimization. A similar result for misspecified linear models was given in Kong and Wiens;[8] in both cases, the proofs run along standard lines based on methods of Knight[25] and Koenker.[1] In Theorem 1, the additional complications posed by the nonlinearity are dealt with via first-order expansions. Complete details may be found in the online supplement.[26]

We require the vector and matrices

$$\boldsymbol{\mu} = \sum_{i=1}^{N} \delta^*(\boldsymbol{x}_i)\boldsymbol{f}(\boldsymbol{x}_i, \boldsymbol{\theta}_\tau)\xi_i : p \times 1$$

$$\boldsymbol{P}_0 = \sum_{i=1}^{N} \boldsymbol{f}(\boldsymbol{x}_i, \boldsymbol{\theta}_\tau)\boldsymbol{f}'(\boldsymbol{x}_i, \boldsymbol{\theta}_\tau)\xi_i : p \times p$$

$$\boldsymbol{P}_1 = \sum_{i=1}^{N} \boldsymbol{f}(\boldsymbol{x}_i, \boldsymbol{\theta}_\tau)\boldsymbol{f}'(\boldsymbol{x}_i, \boldsymbol{\theta}_\tau)\frac{\xi_i}{\sigma(\boldsymbol{x}_i)} : p \times p$$

$$\boldsymbol{P}_2 = \sum_{i=1}^{N} \boldsymbol{f}(\boldsymbol{x}_i, \boldsymbol{\theta}_\tau)\boldsymbol{f}'(\boldsymbol{x}_i, \boldsymbol{\theta}_\tau)\left(\frac{\xi_i}{\sigma(\boldsymbol{x}_i)}\right)^2 : p \times p$$

The vector $\boldsymbol{\mu}$ appears, in the form $\boldsymbol{P}_1^{-1}\boldsymbol{\mu}$, as the asymptotic bias of the quantile estimate. The matrices $\boldsymbol{P}_0$ and $\boldsymbol{P}_1$ determine the 'sandwich' covariance structure, and $\boldsymbol{P}_2$ appears in the maximum bias.

Theorem 1. Assume that the support of $\boldsymbol{\xi}$ is sufficiently large that $\boldsymbol{P}_0$ and $\boldsymbol{P}_1$ are positive definite. Then the τth quantile regression estimate $\hat{\boldsymbol{\theta}}_{n\tau}$ defined by equation (3) is asymptotically normally distributed:

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

We now introduce a measure of the asymptotic loss when the conditional quantile $Y_\tau(\boldsymbol{x}) = F(\boldsymbol{x}; \boldsymbol{\theta}_\tau) + \delta(\boldsymbol{x})$, for $\boldsymbol{x} \in \chi$, is incorrectly estimated by $\hat{Y}_n(\boldsymbol{x}) = F\big(\boldsymbol{x}; \hat{\boldsymbol{\theta}}_{n\tau}\big)$. This measure is the limiting average mean-squared prediction error

$$\mathrm{amse} = \lim_{\mathrm{n}} \frac{1}{\mathrm{N}} \sum_{\mathrm{i}=1}^{\mathrm{N}} \mathrm{E}\left[\left\{\sqrt{\mathrm{n}}\left(\hat{Y}_{\mathrm{n}}(\boldsymbol{x}_{\mathrm{i}}) - \mathrm{Y}_\tau(\boldsymbol{x}_{\mathrm{i}})\right)\right\}^2\right]$$

Define matrices $\boldsymbol{A} = N^{-1}\sum_{i=1}^{N} \boldsymbol{f}(\boldsymbol{x}_i, \boldsymbol{\theta}_\tau)\boldsymbol{f}'(\boldsymbol{x}_i, \boldsymbol{\theta}_\tau)$ and $\boldsymbol{T}_k = \boldsymbol{P}_1^{-1}\boldsymbol{P}_k\boldsymbol{P}_1^{-1}$, $k = 0, 2$. Let $\boldsymbol{\sigma} \stackrel{def}{=} \big(\sigma(\boldsymbol{x}_1), \ldots, \sigma(\boldsymbol{x}_N)\big)'$. Calculations similar to those in Kong and Wiens,[8] but modified in light of the possible heteroscedasticity, yield that the maximum amse, over model errors δ satisfying (5) and (2i), is $\left\{\frac{\tau(1-\tau)}{g_\varepsilon^2(0)} + \eta^2\right\}$ times

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

where $\nu = \eta^2 / \left\{\frac{\tau(1-\tau)}{g_\varepsilon^2(0)} + \eta^2\right\} \in [0, 1]$. The term $tr(\boldsymbol{A}\boldsymbol{T}_0)$ arises solely from variation, and $ch_{\max}(\boldsymbol{A}\boldsymbol{T}_2)$ – the maximum eigenvalue of $\boldsymbol{A}\boldsymbol{T}_2$ – arises solely from bias. The experimenter is now free to choose ν according to the relative emphasis he/she wishes to place on these two components – neither η, τ nor $g_\varepsilon^2(0)$ needs to be specified.

It is our intention to demonstrate that one can obtain efficient and robust designs $\boldsymbol{\xi}$ by adopting a *minimax* approach, i.e. we aim to minimize the maximum amse through a design

$$\boldsymbol{\xi}_* = \text{argmin}\mathcal{L}_\nu(\boldsymbol{\xi}|\boldsymbol{\theta}_\tau, \boldsymbol{\sigma})$$

Remark : Standard asymptotic theory shows that the asymptotic normality asserted in Theorem 1 continues to hold if the $\sigma(\boldsymbol{x})$ are replaced by consistent estimates $\hat{\sigma}(\boldsymbol{x})$ in $\boldsymbol{P}_1$. As well, if $ch_{\max}(\boldsymbol{AT}_2)$ is simple, so that it is a continuous function of the elements of $\boldsymbol{T}_2$,[27] then $\mathcal{L}_\nu(\boldsymbol{\xi}|\boldsymbol{\theta}_\tau, \hat{\boldsymbol{\sigma}})$ is consistent for $\mathcal{L}_\nu(\boldsymbol{\xi}|\boldsymbol{\theta}_\tau, \boldsymbol{\sigma})$. In our adaptive approach described below, $\sigma(\boldsymbol{x})$ will be replaced by an estimate each time a new design point is to be chosen. That the maximum eigenvalue is simple is checked numerically.

3 Computations and case study

As with all problems in nonlinear design, the dependence of the loss on the unknown parameters is problematic. We shall deal with this in two ways. In the first *sequential*, local design, we assume that the parameter $\boldsymbol{\theta}_\tau$ and the scale function are given, and we construct a design using sequential methods that parallel the development in Wiens.[14] We note that the assumption of a known $\boldsymbol{\theta}_\tau$ is realistic in Phase 2 clinical trials[28] or if a reliable preliminary estimate is available. Here, we shall use the sequential approach primarily to motivate an *adaptive* approach, in which data are also gathered, and the estimates computed and updated, as the experiment progresses.

The methods will be illustrated in the context of a study detailed in Cressie and Keightley,[19] using data obtained from several estrogen receptor assays of human uterine cytosol fraction. Human neuronal development and reorganization are determined by steroid hormones that interact with receptors.[29] Estrogen is the major sex steroid in females, is naturally produced by the ovaries in premenopausal women and its deficiency is a factor for menopause, cardiovascular disease, type 2 diabetes, obesity and hyperlipidemia.[30] As well, low estrogen levels are related to mood disorders and depression in women.[31] Treatments are available to balance estrogen level, thus reducing the risk of cardiovascular disease and mortality.[32]

Cressie and Keightley[19] collected the amount of estrogen (y) bound to a receptor and amount of hormone (x) and fit Michaelis–Menten response models using least squares to estimate the conditional means. Their data, along with the estimated response curves with parameters ν, are shown in Figure 1. Also shown are several τ-quantile curves, obtained from the simulated data of Section 3.2.

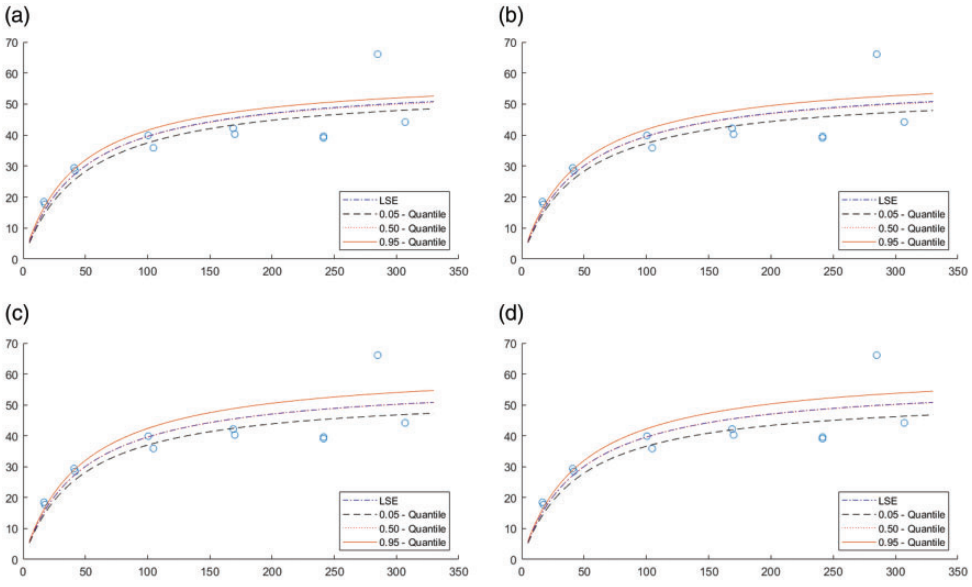

Figure 1. Data gathered by Cressie and Keightley[19] with the least squares response curve and various τ-quantile curves: (a) $\nu = .20$, (b) $\nu = .35$, (c) $\nu = .75$, (d) $\nu = .95$.

In this section, we describe our sequential local (SL; 'SL designs') and adaptive (A; 'Adaptive designs') design methods to construct designs for a future experiment, with the aim of estimating and predicting conditional quantiles. As in Cressie and Keightley,[19] the model being fitted is assumed to have a Michaelis–Menten response $F(x;\boldsymbol{\theta}_\tau) = \theta_{\tau 1}x/(\theta_{\tau 2}+x)$, although the experimenter seeks robustness against misspecifications as described in Section 2. The design space partitions $[5,400]$: $\chi = \{x_i = 5 + 395(i-1)/(N-1);\ i = 1,\cdots,N\}$ for $N = 100$. The computing code, in Matlab, is available from us.

3.1 SL designs

In this section, we design for local parameters $\boldsymbol{\theta}_0 = (57.98, 46.43)'$ obtained as estimates by Cressie and Keightley.[19] We assume known scales $\sigma_0(x)$ proportional to one of $\{1, x+x_N, \sqrt{x+x_N}, x+0.25x_N\}$, normalized by (2ii). We begin with a two-point design $\xi_2 = (.5, 0, \ldots, 0, .5)'$ placing one of the two observations at each endpoint of χ. For $n = 2, 3, \ldots, 999$, given $\boldsymbol{\xi}_n$, we set $\boldsymbol{\xi}_{n+1} = \boldsymbol{\xi}_{n+1,i^*}$, where, with

$$\boldsymbol{\xi}_{n+1,i} \stackrel{def}{=} \frac{n\boldsymbol{\xi}_n + \left(0,\ldots,0,1^{\downarrow i},0,\ldots,0\right)'}{n+1}$$

and $\boldsymbol{\sigma}_0 \stackrel{def}{=} \left(\sigma_0(x_1),\ldots,\sigma_0(x_N)\right)'$, we choose

$$i^* = \arg\min_{i\in\{1,\ldots,N\}} \mathcal{L}_\nu\left(\boldsymbol{\xi}_{n+1,i}|\boldsymbol{\theta}_0,\boldsymbol{\sigma}_0\right) \tag{6}$$

A study of similar methods by Wiens[14] suggests that the designs obtained in this manner will be at least very close to optimal in the class of all designs. We stopped at a design size of 1000 since there seemed very little to be gained by continuing. In anticipation of the experimenter seeking a final design of size 30, we then chose design points

$$x_{(i)} = \boldsymbol{\xi}_{1000}^{-1}\left(\frac{i-.5}{30}\right), i = 1,\ldots,30 \tag{7}$$

Note that the designs become markedly more diffuse with increasing ν, reflecting the increasing emphasis on bias (see Figure 2).

3.2 Adaptive designs

Our adaptive method parallels the sequential method, but with equation (6) replaced by

$$i^* = \arg\min_{i\in\{1,\ldots,N\}} \mathcal{L}_\nu\left(\boldsymbol{\xi}_{n+1,i}|\hat{\boldsymbol{\theta}}_{n\tau},\hat{\boldsymbol{\sigma}}_n\right) \tag{8}$$

Here, $\hat{\boldsymbol{\theta}}_{n\tau}$ is computed from $\{x_(i), y_i\}_{i=1}^n$ and $\hat{\boldsymbol{\sigma}}_n$ is the N-vector of estimates $\hat{\sigma}_n(x_i)$. To compute $\hat{\boldsymbol{\sigma}}_n$, we first obtain Gaussian kernel estimates

$$s_n^2(x) = \sum_{i=1}^n w(x-x_{(i)})(Y_i - F(x_{(i)};\hat{\boldsymbol{\theta}}_{n\tau}))^2$$

where $w(t_i) = \exp\left\{-\frac{1}{2}\left(\frac{t_i}{h}\right)^2\right\}/\sum_{i=1}^n \exp\left\{-\frac{1}{2}\left(\frac{t_i}{h}\right)^2\right\}$, with 'bandwidth' $h = 40$. See Remark (1) at the end of this section, for a description of our method of choosing the bandwidth. We then fit a least squares regression line $\hat{\beta}_0 + \hat{\beta}_1 x$ to the data $\{s_n(x_{(i)})\}_{i=1}^n$ and take $\hat{\sigma}_n(x_i) = \hat{\beta}_0 + \hat{\beta}_1 x_i$. These estimates are then normalized as at (2ii). If the standard deviations really are linear in x, then (under conditions on $h = h_n$) $\hat{\boldsymbol{\sigma}}_n$ is consistent; more importantly we have found this to be an effective method of choosing $x_{(n+1)}$ even when the scale function is misspecified.

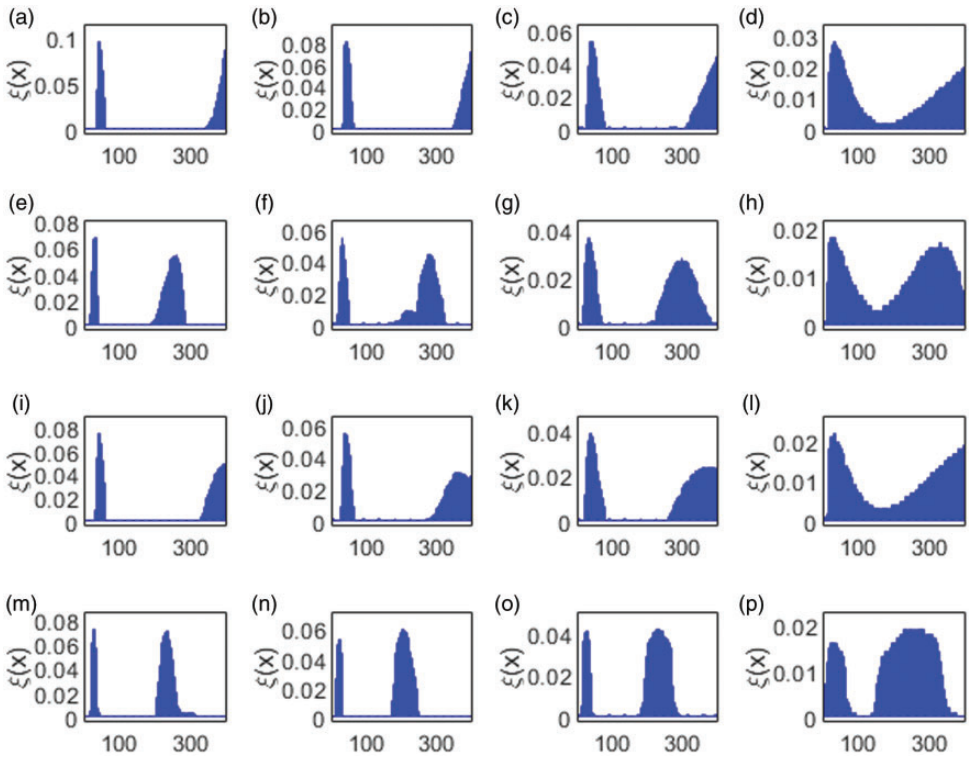

Figure 2. One thousand-point SL design, with 30-point designs marked with red bullet points. Rows 1–4 use $\sigma(x) \propto 1, x + x_N, \sqrt{x + x_N}, x + 0.25x_N$, respectively; columns 1–4 use $\nu = 0.20, 0.35, 0.65, 0.95$, respectively.

We illustrate the method through a simulation study, continuing Cressie and Keightley.[19] Simulated data are generated by

$$Y(x) = F(x; \boldsymbol{\theta}_\tau) + \delta(x) + \sigma_0(x)\varepsilon$$

where $\boldsymbol{\theta}_\tau = (57.98, 46.43)'$, $\varepsilon \sim N\{-\sigma_\varepsilon \Phi^{-1}(\tau), \sigma_\varepsilon^2\}$, $\sigma_\varepsilon = 2, \Phi(\cdot)$ is the standard normal distribution function, $\sigma_0(x)$ is any of the scale functions in §3.1 and $\delta(x) = \sigma_0(x)\delta^*(x)$ with $\boldsymbol{\delta}^* = \left(\delta^*(x_1), \ldots, \delta^*(x_N)\right)'$ determined as follows.

Let $\boldsymbol{F}$ be the $N \times p$ matrix with rows $\{\boldsymbol{f}'(x; \boldsymbol{\theta}_\tau) | x \in \chi\}$ and let

$$\boldsymbol{F} = \left(\boldsymbol{Q}_1 \vdots \boldsymbol{Q}_2\right) \begin{pmatrix} \boldsymbol{R}_{p\times p} \\ \boldsymbol{0}_{N-p\times p} \end{pmatrix}$$

be the qr-decomposition, where $\left(\boldsymbol{Q}_1 \vdots \boldsymbol{Q}_2\right)$ is an $N \times N$ orthogonal matrix and $\boldsymbol{R}$ is triangular. Condition (5) states that $\boldsymbol{\delta}^*$ is orthogonal to the column space of $\boldsymbol{F}$, hence to that of $\boldsymbol{Q}_1$, and so must lie in the column space of $\boldsymbol{Q}_2$. Along with (2i), this forces

$$\boldsymbol{\delta}^* = \eta \sqrt{\frac{N}{n}} \boldsymbol{Q}_2 \boldsymbol{t}_{N-p\times 1}$$

where $||\boldsymbol{t}|| \leq 1$. In our simulations, we took $\eta = 1$, $\tau = .5$ and $\boldsymbol{t} = \boldsymbol{z}/||\boldsymbol{z}||$, with $\boldsymbol{z} \sim N(\boldsymbol{0}, \boldsymbol{I}_{N-p})$. Then, for $\nu \in \{0.20, 0.35, 0.65, 0.95\}$ and a desired study size of n_*, we carry out the following.

3.2.1 *Initialization step*

Start with a design $\boldsymbol{\xi}_{n_0}$ with mass $1/n_0$ on each of n_0 equally spaced points. Determine $\hat{\boldsymbol{\theta}}_{n_0\tau}$ from equation (3), for this, we use the Matlab minimizer *fminsearch* and compute $\hat{\boldsymbol{\sigma}}_{n_0}$ as described above.

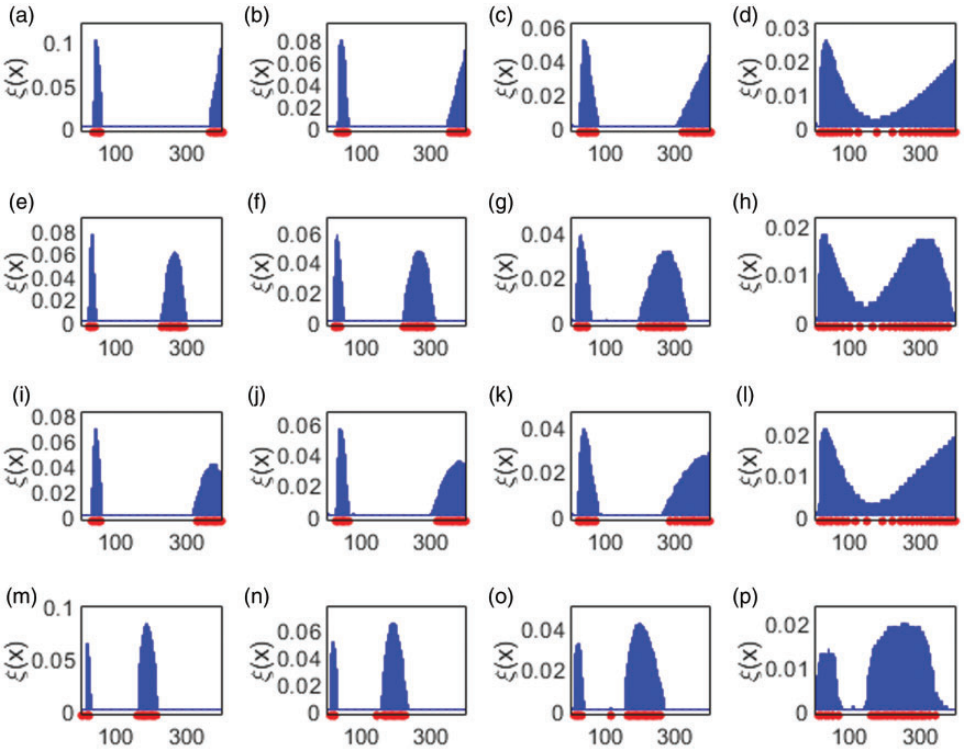

Figure 3. Adaptive designs on 1000 points. Rows and columns are as in Figure 2.

3.2.2 *Adaptive step*

For $n = n_0, \ldots, n_* - 1$, determine $\boldsymbol{\xi}_{n+1}$ from (8) and draw a further observation $y_{n+1} = F(x_{i_*}; \boldsymbol{\theta}_\tau) + \delta(x_{i_*}) + \sigma_0(x_{i_*})\varepsilon$.

For comparison with the sequential approach, we first carried out this procedure with $n_0 = 10$, $n_* = 1000$. The resulting designs are shown in Figure 3, and the scale estimates are shown in Figure 4. The designs are very close to those obtained sequentially, and the scale estimates are generally very accurate. Figure 6 gives the comparative ratios of the minimized losses, relative to those of the adaptive design. From these, we see that the adaptive designs perform as well as the sequential designs when the scale function is correctly specified, better when the sequential designs are based on a misspecified scale function.

Does the very favourable asymptotic performance of the adaptive designs carry over to smaller, more realistic study sizes? To answer this, we computed them with $n_0 = 2$, $n_* = 30$. The initial design had $\boldsymbol{\xi}_2(x_2) = \boldsymbol{\xi}_2(x_{75}) = .5$. We carried out this procedure $M = 100$ times, using a different $\boldsymbol{\delta}_m^*$ each time. This yielded designs $\{x_{(i)m}\}_{i=1}^{30}$ for $m = 1, \ldots, M$. In Figure 5, we present representative realizations of these designs. Then, to limit the simulation variability, we present, in Figure 7, the average losses resulting from these M designs, compared to those of the 30-point sequential designs that were exhibited in Figure 2. While the comparisons with the sequential designs are not as striking as those in Figure 6, the adaptive designs still perform very favourably – especially so since the sequential designs are based on prior knowledge that is almost never available to the experimenter. Those shown in Figure 5 are typical in placing mass more uniformly throughout χ than do their sequential counterparts, while still emphasizing those regions in which the 30-point sequential designs are concentrated. Only rarely does the adaptive approach call for replication – something that is precluded by equation (7) in the sequential approach.

Remarks

1. The choice of the bandwidth in kernel estimation is typically a somewhat thorny issue, guided both by theory and by ad hoc considerations. A complicating factor in design is that the optimal designs tend to be far from uniform (a typical, asymptotic requirement for optimality of kernel methods) so that if the bandwidth is too small, then, often, the 'windows' will be essentially empty. In this study, we chose h as follows. After obtaining the initial sample of size $n_0 = 10$, we computed and plotted kernel estimates $s_n(x)$ at each of these points, using

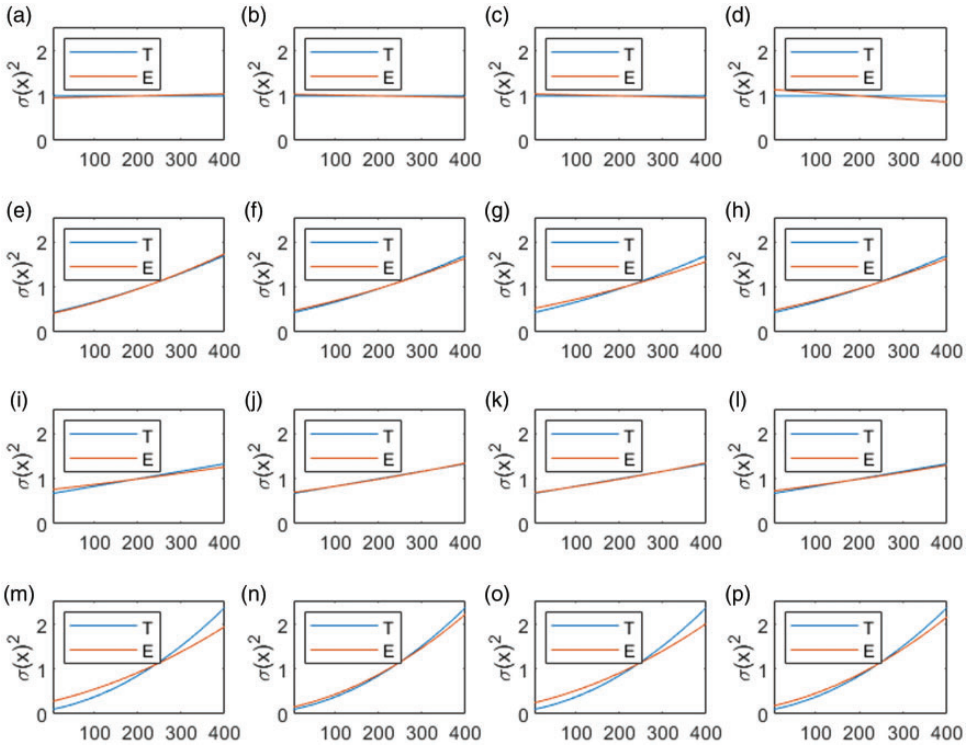

Figure 4. True (T) and estimated (E) scale functions resulting from the 1000-point adaptive designs. Rows and columns are as in Figure 2.

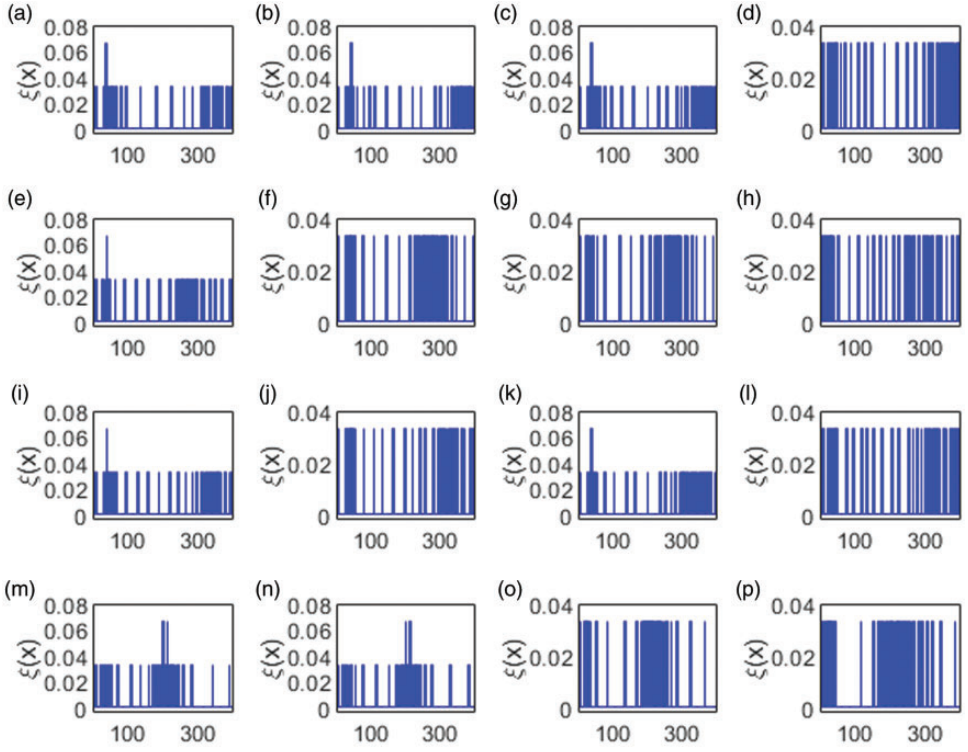

Figure 5. Typical 30-point adaptive designs. Rows and columns are as in Figure 2.

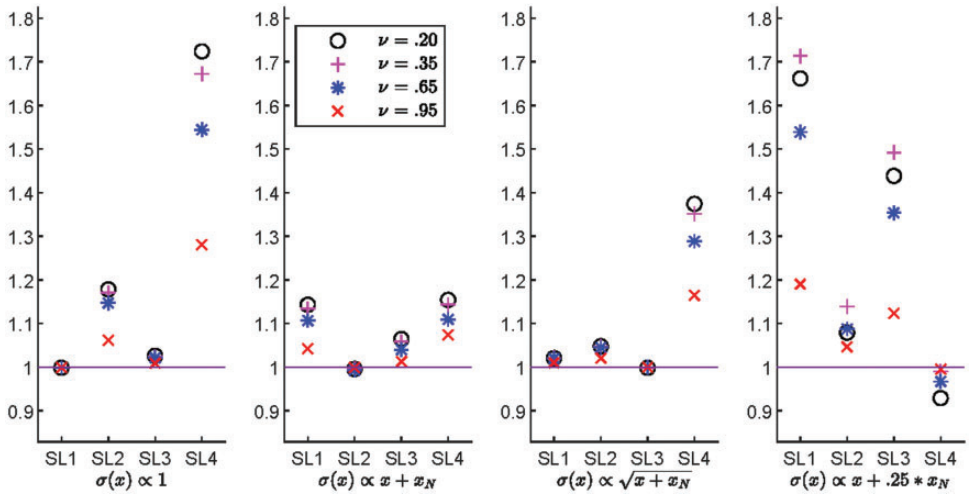

Figure 6. Losses arising from 1000-point designs relative to those of the adaptive designs, for the given scale functions $\sigma(x)$. Designs SL1–SL4 assume $\sigma(x) \propto 1, x + x_N, \sqrt{x + x_N}, x + 0.25x_N$, respectively.

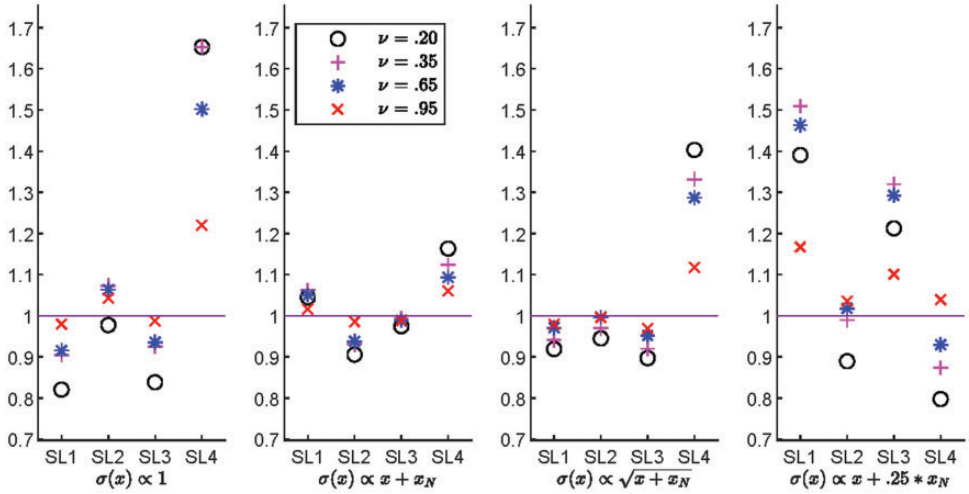

Figure 7. Losses arising from 30-point designs relative to those of the adaptive designs, for the given scale functions $\sigma(x)$. Designs SL1–SL4 assume $\sigma(x) \propto 1, x + x_N, \sqrt{x + x_N}, x + 0.25x_N$, respectively.

a range of bandwidths from very small to very large. As expected, bandwidths which were much too small resulted in seemingly random, highly varied estimates. Bandwidths which were too large resulted in estimates with very little variation. Near our chosen value of $h = 40$, the estimates stabilized and began to exhibit (and retain) some x-dependent structure. It is perhaps not an accident that this choice of h ensures that each of the equally spaced initial points will have its two nearest neighbours within one standard deviation of the Gaussian numerator in the kernel estimate so that they will make a reasonable contribution to the estimate at that point.

2. In other simulations, not shown here, we found that the final designs were fairly insensitive to the choice of bandwidth, over a reasonable range.
3. We have simulated data, using the least squares estimates obtained by Cressie and Keightley,[19] only for the purposes of comparison with their estimates (see Figure 1). The close agreement between their estimates, and ours when $\tau = .5$, would seem to support their assumption of a symmetric error distribution. We also used least squares as the starting value for the Matlab minimizer used in the initialization step of our adaptive algorithm; at all subsequent steps, the starting value was the final value obtained before the addition of a subsequent design point. This is in line with comments of Huber[33] who pointed out that in regression, as opposed to location estimation, there is no easy analogue of the, very robust, sample median – one generally must resort to least squares for starting values.

4. Note that we are basing our methods in this section on the asymptotic normality asserted in Theorem 1, which is rigorous only for independent observations. But previous studies of adaptive design – see for instance Ford and Silvey,[9] Chaudhuri and Mykland[34] and Wu[10] – have repeatedly made the point also emphasized by Sinha and Wiens,[13] who state 'We conclude that estimates computed after the experiment has been carried out [adaptively] may, in only moderately sized samples, be safely used to make standard normal-theory inferences, ignoring the dependencies arising from the [adaptive] nature of the sampling'.

Acknowledgements

This work has benefited greatly from the comments of several anonymous referees.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: Natural Sciences and Engineering Council of Canada and Canadian Statistical Sciences Institute.

ORCID iDs

Selvakkadunko Selvaratnam https://orcid.org/0000-0001-9271-1827
Douglas P Wiens https://orcid.org/0000-0002-0439-4145

References

1. Koenker R. *Quantile regression*. New York: Cambridge University Press, 2005.
2. Koenker R. Quantile regression: 40 years on. *Annu Rev Econom* 2017; **9**: 155–176.
3. Yu K, Lu Z and Stander J. Quantile regression: applications and current research areas. *J R Stat Soc Ser D* 2003; **52**: 331–350.
4. Geraci M. Modelling and estimation of nonlinear quantile regression with clustered data. *Comput Stat Data Anal* 2019; **136**: 30–461.
5. Farcomeni A and Geraci M. Multistate quantile regression models. *Stat Med* 2020; **39**: 45–56.
6. Wiens DP. *Robustness of design*. In: Dean A, Morris M, Stufken J and Bingham D (eds) *Handbook of design and analysis of experiments*. Boca Raton, FL: Chapman & Hall/CRC, 2015.
7. Dette H and Trampisch M. Optimal designs for quantile regression models. *J Am Stat Assoc* 2012; **107**: 1140–1151.
8. Kong L and Wiens DP. Model-robust designs for quantile regression. *J Am Stat Assoc* 2015; **110**: 233–245.
9. Ford I and Silvey SD. A sequentially constructed design for estimating a nonlinear parametric function. *Biometrika* 1980; **67**: 381–388.
10. Wu CFJ. Asymptotic inference from sequential design in a nonlinear situation. *Biometrika* 1985; **72**: 553–558.
11. Santos MI and Santos PM. Sequential experimental designs for nonlinear regression metamodels in simulation. *Simul Model Pract Theory* 2008; **16**: 1365–1378.
12. Guest T and Curtis A. Iteratively constructive sequential design of experiments and surveys with nonlinear parameter-data relationships. *J Geophys Res-Solid Earth* 2009; **114**: 1–14.
13. Sinha S and Wiens DP. Robust sequential designs for nonlinear regression. *Canad J Stat* 2002; **30**: 601–618.
14. Wiens DP. I-robust and D-robust designs on a finite design space. *Stat Comput* 2018; **28**: 241–258.
15. Müller H-H and Schäfer H. A general statistical principle for changing a design any time during the course of a trial. *Stat Med* 2004; **23**: 2497–2508.
16. Konig F. Multiplicity and flexibility in clinical trials. *Pharm Stat* 2007; **6**: 205–216.
17. Levin GP, Emerson SS and Emerson SC. Adaptive clinical trial designs with pre-specified rules for modifying the sample size: understanding efficient types of adaptation. *Stat Med* 2013; **13**: 1259–1275.
18. Magirr D, Stallard N and Jaki T. Flexible sequential designs for multi-arm clinical trials. *Stat Med* 2014; **33**: 3269–3279.
19. Cressie NAC and Keightley DD. The underlying structure of the direct linear plot with application to the analysis of hormone-receptor interactions. *J Steroid Biochem* 1979; **11**: 1173–1180.
20. Jurečková J and Prochazka B. Regression quantiles and trimmed least squares estimators in nonlinear regression models. *J Nonparam Stat* 1994; **3**: 201–222.
21. Oberhofer W and Haupt H. Asymptotic theory for nonlinear quantile regression under weak dependence. *Econom Theory* 2016; **32**: 686–713.

22. Kim TH and White H. Estimation, inference, and specification testing for possibly misspecified quantile regressions. *Adv Econometr* 2003; **17**: 107–132.
23. Komunjer I. Quasi-maximum likelihood estimation for conditional quantiles. *J Econometr* 2005; **128**: 137–164.
24. Angrist J, Chernozhukov V and Fernández-Val I. Quantile regression under misspecification, with an application to the U. S. wage structure. *Econometrica* 2006; **74**: 539–563.
25. Knight K. Limiting distributions for L_1 regression estimators under general conditions. *Ann Stat* 1998; **26**: 755–770.
26. Selvaratnam S. A central limit theorem for nonlinear quantile regression. https://sites.ualberta.ca/∼dwiens/home page/techReports/TR S139.pdf (2020, accessed 23 July 2020).
27. Magnus JR. On differentiating eigenvalues and eigenvectors. *Econom Theory* 1985; **1**: 179–191.
28. Lange MR and Schmidli H. Optimal design of clinical trials with biologics using dose-time-response models. *Stat Med* 2014; **33**: 5249–5264.
29. Kawata M. Roles of steroid hormones and their receptors in structural organization in the nervous system. *Neurosci Res* 1995; **24**: 1–46.
30. Ainslie DA, Morris MJ, Wittert G, et al. Estrogen deficiency causes central leptin insensitivity and increased hypothalamic neuropeptide Y. *Int J Obesity* 2001; **25**: 1680–1688.
31. Wharton W, Gleason CE, Olson SR, et al. Neurobiological underpinnings of the estrogen – mood relationship. *Curr Psychiatr Rev* 2012; **8**: 247–256.
32. Sood R, Faubion SS, Kuhle CL, et al. Prescribing menopausal hormone therapy: an evidence-based approach. *Int J Women's Health* 2014; **6**: 47–57.
33. Huber P. Robust regression: asymptotics, conjectures and Monte Carlo. *Ann Stat* 1973; **1**: 799–821.
34. Chaudhuri P and Mykland P. Nonlinear experiments: optimal design and inference based on likelihood. *J Am Stat Assoc* 1993; **88**: 538–546.
35. Sinha S and Wiens DP. Asymptotics for robust sequential designs in misspecified regression models. In: Moore M, Léger C and Froda S (eds) *IMS lecture notes – monograph series*. Mathematical Statistics and Applications: Festschrift for Constance van Eeden, 2003, pp.223–248.