Robust Designs for Haar Wavelet Approximation Models

Xiaojian Xu[1], **Lin Zhao**[2]

Department of Mathematics, Brock University
St. Catharines, Ontario, T6G 2G1, Canada
[1] e-mail: `xxu@brocku.ca`; [2] e-mail: `lz07tu@badger.ac.brocku.ca`

March 5, 2010

Abstract In this article, we discuss the construction of robust designs for heteroscedastic wavelet regression models when the assumed models are possibly contaminated over two different neighbourhoods: G_1, and G_2. Our main findings are: (1) A recursive formula for constructing D-optimal designs under G_1; (2) Equivalency of Q-optimal and A-optimal designs under both G_1 and G_2; (3) D-optimal robust designs under G_2; and (4) Analytic forms for A- and Q-optimal robust design densities under G_2. Several examples are given for the comparison, and the results demonstrate that our designs are efficient.

Key words A-optimal, D-optimal, Q-optimal, Heteroscedasticity, Minimax, Design Discretization.

1 Introduction

This paper aims to investigate the optimal designs of experiments for wavelet regression models. The designs constructed are robust against the approximation in the assumed finite terms of wavelets. Minimax approach is adapted to address the worst possible situation within a contamination class. Two different contamination classes are discussed in this paper.

Wavelet regression models are widely used in many areas, such as spectroscopy (Brown, Fearn, and Vannucci, 2001), signal processing (Robert and Richard, 1999), and feature detection in earth science (Murat and Necati, 2006). The development of wavelet theory has also extended its applications in fast algorithms for integral transforms in numerical analysis (see Alpert, 1992). Another application is in function representation methods, and such application has stimulated interest in wavelet approximations of regression response functions for the analysis of experimental data (see Oyet and Wiens (2000) and the references cited in).

Sprang (1989) discussed Haar's wavelet and indicated that Haar's wavelet, as a piecewise constant example, "reveals a lot with no deep analysis" (p. 615). The reasons we utilize the Haar's wavelet in this paper are: First, piecewise linear regression is the simplest in terms of implementation and is often used with acceptable accuracy; and second, the function that we frequently encounter may not even be continuous. Capilla (2005) also gave the real world examples of Haar wavelet application in detecting microseismic signal arrivals

Optimal design for wavelet regression is a relatively new topic developed in the field of statistics. Some efficient designs for certain specific wavelet regression models have been investigated by Herzberg and Traves (1994), Xie (2002), and Tian and Herzberg (2006A), among others. Herzberg and Traves (1994) discussed D-optimal designs for the Haar wavelet model and Xie (2002) provided the D-optimal designs for b-adic Haar wavelet models; more recently, Tian and Herzberg (2006A) constructed D-optimal designs for a combined linear and Haar wavelet function.

Traditionally, an optimal design is chosen in order that the covariance matrix of the estimates can be minimized, resulting in the minimum of the covariance matrix means the minimum of an appropriate real-valued function of the covariance matrix. There are three major criteria for optimal designs, namely D-optimal, A-optimal and Q-optimal, which are to minimize the determinant, the trace and the average prediction of covariance matrix, and it has been shown that an optimal design is efficient to estimate the parameters if the assumed model is correct. However, Box and Draper (1959) exposed the risks of designing a regression experiment which assumes the fitted model is exactly correct; they found that very small deviations from an assumed model can eliminate any supposed gains arising from the use of a design which minimizes variance alone. Tian and Herzberg (2006B) considered a Haar wavelet approximation model with heteroscedasticity, and obtained the A-minimax designs over an L_1 type of contamination class; yet, they left a rather open problem on obtaining D-minimax designs. Oyet and Wiens (2000) considered a wavelet approximation model but over an L_2 type of contamination class, and obtained continuous A-, D-, and Q-minimax designs for the model with homoscedasticity.

In this paper,

(1). We first tackle the unsolved problem pointed out in Tian and Herzberg (2006B): D- and Q- minimax design construction for a Haar wavelet approximation model with a possible contaminant which is a member of an L_1 type of contamination class.

(2). Then, we discuss the D-, A-, and Q-optimal robust designs under a much broader L_2 type of contamination class. We extend the work of Oyet and Wiens (2000) on homoscedastic wavelet models to heteroscedastic wavelet models.

The rest of this article is organized as follow: some mathematical preliminaries are detailed in §2; the designs for (1) are provided in §3; the designs for (2) are given in §4; design implementation and comparison of the robust designs obtained in this paper to commonly used uniform designs, as well as a simulation study, are demonstrated with several examples in §5; and concluding remarks are presented in §6. All derivations are provided in an appendix.

2 Preliminaries and Notation

A wavelet is a mathematical function used to divide a given function or continuous-time signal into different scale components. As Tian and Herzberg (2006B) indicated, a wavelet system can be considered as a basis for representing square integrable functions in different scales, in a similar way to that of polynomials or trigonometric functions. Haar (1910) discovered a wavelet system on the real line, and its orthogonal basis in $L_2(\mathbb{R})$ is generated by the Haar scaling function $\phi(x)$ and the Haar primary wavelet $\psi(x)$, where

$$\phi(x) = \begin{cases} 1 & (0 \leq x < 1) \\ 0 & \text{otherwise}, \end{cases}$$

and

$$\psi(x) = \begin{cases} 1 & \left(0 \leq x < \frac{1}{2}\right) \\ -1 & \left(\frac{1}{2} \leq x < 1\right) \\ 0 & \text{otherwise}. \end{cases}$$

Haar wavelets are the simplest wavelets: they only have three possible values on the real line $\mathbb{R}$ and each of them is a step function. For reader's convenience, we also keep the same notions in both this and next section as used in Tian and Herzberg (2006B). Herzberg and Traves (1994) used a Haar wavelet function to fit the model. Under the least squares estimation, the authors proved the following result: for the Haar wavelet model of order m

$$E(y(x)) = \beta_0 + \sum_{j=0}^{m} \sum_{k=0}^{2^j-1} \beta_{jk} \psi_{jk}(x), \tag{1}$$

with

$$\psi_{jk}(x) = 2^{\frac{j}{2}}\psi\left(2^j x - k\right) = \begin{cases} 2^{\frac{j}{2}} & \text{if } \frac{k}{2^j} \le x < \frac{k}{2^j} + \frac{1}{2^{j+1}}, \\ -2^{\frac{j}{2}} & \text{if } \frac{k}{2^j} + \frac{1}{2^{j+1}} \le x < \frac{k}{2^j} + \frac{1}{2^j}, \\ 0 & \text{otherwise,} \end{cases}$$

as Haar wavelets with $j = 0, \cdots m$; $k = 0, \cdots, 2^j - 1$, any design which concentrates mass $2^{-(m+1)}$ in the 2^{m+1} intervals

$$\left\{\left[2^{-(m+1)}i, 2^{-(m+1)}(i+1)\right)\right\}_{i=0,1,\cdots,2^{m+1}-1},$$

is D-optimal design, where β_0, β_{jk} $\left(j = 0, \cdots m;\ k = 0, \cdots, 2^j - 1\right)$ are unknown parameters and are to be estimated by the least squares method.

As in Daubechies (1992), we approximate the regression response by finitely many dominant terms of its wavelet representation, with remainder $g(x)$. The true model may be written as

$$y(x) = \beta_0 + \sum_{j=0}^{m}\sum_{k=0}^{2^j-1}\beta_{jk}\psi_{jk}(x) + g(x) + \varepsilon(x), \tag{2}$$

with

$$\varepsilon(x) \in \Upsilon = \left\{E\{\varepsilon(x)\} = 0,\quad Var\{\varepsilon(x)\} = \sigma^2\pi^{-1}(x),\ \ Cov\{\varepsilon(x), \varepsilon(x')\} = 0 \ \text{ for } x \neq x'\right\}, \tag{3}$$

where $\pi(x)$ is a known efficiency function, and $g(x)$ is an unknown contaminant to the fitted model (1). The fitted response (1) is typically acknowledged to be only an approximation for a x belonging to a bounded design space S. The least squares estimates $\hat{\boldsymbol{\beta}}$ of $\boldsymbol{\beta} = \left(\beta_0, \beta_{00}, \beta_{10}, ..., \beta_{m0}, ..., \beta_{m,2^m-1}\right)^T$ and $\hat{Y} = \hat{\beta}_0 + \sum_{j=0}^{m}\sum_{k=0}^{2^j-1}\hat{\beta}_{jk}\psi_{jk}(x)$ of $E(Y|\mathbf{x})$ are possibly biased if the response is misspecified. In this situation, robust designs can play an important role in choosing optimal design points x_1,..., $x_n \in S$ so that estimates $\hat{\boldsymbol{\beta}}$ and $\hat{Y}$ remain relatively efficient, with a small bias caused by the model misspecification.

In the next three sections, we will present solutions to (1) and (2) stated in §1.

3 Robust Designs for the Approximate Haar wavelet Models over G_1

In order to control the magnitude on the bias, as considered in Tian and Herzberg (2006B), the contamination function $g(x)$ in this section is assumed to be an unknown member of contamination class G_1 where

$$G_1 = \left\{g(x) \,|\, |g(x)| \le \tau,\ \int_0^1 g(x)\,dx = 0 \text{ and } \int_0^1 \psi_{jk}(x)\,g(x)\,dx = 0\right\}. \tag{4}$$

The first condition in G_1 is a magnitude restriction on $g(x)$. The other two conditions are requirements of orthogonality.

We consider a Haar wavelet approximation model (2), where $\varepsilon(x) \in \Upsilon$, and $g(x) \in G_1$. Tian and Herzberg (2006B) have shown that the mean squared error matrix for the ordinary least squares estimator is a function of contaminant $g(x)$ and design ξ with support points $(x_1, ..., x_{2^m+1})$ and corresponding allocations $(n_1, ..., n_{2^m+1})$. The mean squared error matrix is

$$\mathbf{M}(g,\xi) = 2^{-2(m+1)}\mathbf{X}_{\mathbf{H}}^T\left(\sigma^2\mathbf{L}^{-1} + \mathbf{g}_B\mathbf{g}_B^T\right)\mathbf{X}_{\mathbf{H}}, \tag{5}$$

where

$$\mathbf{X_H} = \begin{bmatrix} 1 & \psi_{00}(u_1) & \cdots & \psi_{m,2^m-1}(u_1) \\ 1 & \psi_{00}(u_2) & \cdots & \psi_{m,2^m-1}(u_2) \\ \vdots & \vdots & \ddots & \vdots \\ 1 & \psi_{00}(u_{2^{m+1}}) & \cdots & \psi_{m,2^m-1}(u_{2^{m+1}}) \end{bmatrix}_{2^{m+1},2^{m+1}},\ u_i = \frac{(i-1)}{2^{m+1}}, \tag{6}$$

$$\mathbf{g}_B = \begin{bmatrix} \frac{\left[\pi(x_{11})g(x_{11})+\cdots+\pi\left(x_{1n_1}\right)g\left(x_{1n_1}\right)\right]}{l_1} \\ \vdots \\ \frac{\left[\pi\left(x_{2^{m+1},1}\right)g\left(x_{2^{m+1},1}\right)+\cdots+\pi\left(x_{2^{m+1},n_{2^{m+1}}}\right)g\left(x_{2^{m+1},n_{2^{m+1}}}\right)\right]}{l_{2^{m+1}}} \end{bmatrix}, \text{ and} \tag{7}$$

$$\mathbf{L} = \mathbf{diag}\left\{l_1, l_2, \cdots, l_{2^{m+1}}\right\}, \text{ with } l_i = \pi\left(x_{i1}\right) + \cdots + \pi\left(x_{in_i}\right). \tag{8}$$

3.1 D-minimax Designs for the Approximate Haar wavelet Models over G_1

The D-criterion for the approximately linear Haar wavelet model is the determinant of the MSE matrix, namely $\Phi_{\mathbf{D}}\left(g,\xi\right) = \left|\mathbf{M}\left(g,\xi\right)\right|$. By dividing the design space $[0,1)$ into 2^{m+1} subintervals as

$$\chi_i = \left[\frac{i-1}{2^{m+1}}, \frac{i}{2^{m+1}}\right), \ i = 1, \cdots, 2^{m+1},$$

we suppose that the efficiency function $\pi\left(x\right)$ is known, and that the maximum value of $\pi\left(x\right)$ on χ_i is Z_i $\left(i = 1, \cdots, 2^{m+1}\right)$. Then, the maximum of $\Phi_{\mathbf{D}}\left(g,\xi\right)$ over G_1 attains its minimum value when the design support points $x_{ij} = x_{(i)}$ $\left(i = 1, \cdots, 2^{m+1};\ j = 1, \cdots, n_i\right)$, where $x_{(i)}$ are the points such that $\pi\left(x_{(i)}\right) = Z_i$ and $n_1, n_2, \cdots, n_{2^{m+1}}$ are positive real numbers that minimize the maximum of $\Phi_{\mathbf{D}}\left(g,\xi\right)$ over $g \in G_1$, which is:

$$\max_{g\in G_1} \Phi_{\mathbf{D}}\left(g,\xi\right) = \frac{\sigma^{2^{m+2}-2}\tau^2}{\left(2^{m+1}\right)^{2^{m+1}} Z_1 Z_2 \cdots Z_{2^{m+1}}} \frac{\sigma^2\tau^{-2} + n_1 Z_1 + \cdots + n_{2^{m+1}} Z_{2^{m+1}}}{n_1 n_2 \cdots n_{2^{m+1}}}, \tag{9}$$

subject to $n_1 + n_2 + \cdots + n_{2^{m+1}} = n$. Tian and Herzberg (2006B) also indicate that the theoretical solution of $n_1, n_2, \cdots, n_{2^{m+1}}$ is difficult to obtain although the minimum value of $\max_{g\in G_1} \Phi_{\mathbf{D}}\left(g,\xi\right)$ exists. The following section will tackle this open problem of obtaining the optimal allocations $n_i's$. Because of the fact that σ, τ, m, and Z_i are all known positive numbers, our target of obtaining the minimum of $\max_{g\in G_1} \Phi_{\mathbf{D}}\left(g,\boldsymbol{\xi}\right)$ becomes finding the minimum of $\frac{\sigma^2\tau^{-2}+n_1Z_1+\cdots+n_{2^{m+1}}Z_{2^{m+1}}}{n_1 n_2 \cdots n_{2^{m+1}}}$ subject to $n_1 + n_2 + \cdots + n_{2^{m+1}} = n$. We have the following theorem:

Theorem 1 *Suppose that the efficiency function $\pi\left(x\right)$ is known and the maximum value of $\pi\left(x\right)$ on χ_i is Z_i $\left(i = 1, \cdots, 2^{m+1}\right)$. Then the minimum of (9) is attained when the design support points $x_{ij} = x_{(i)}$ $\left(i = 1, \cdots, 2^{m+1}; j = 1, \cdots, n_i\right)$, where $x_{(i)}$ are the points such that $\pi\left(x_{(i)}\right) = Z_i$, and $n_1, n_2, \cdots, n_{2^{m+1}}$ are positive real numbers which satisfy*

$$\frac{n_k}{n} = \frac{\sqrt{\frac{\sigma^2\tau^{-2}}{\prod_{j=1,j\neq k}^{2^{m+1}} n_j} + \sum_{i=1,i\neq k}^{2^{m+1}} \frac{Z_i}{\prod_{j=1,j\neq i,j\neq k}^{2^{m+1}} n_j}}}{\sum_{k=1}^{2^{m+1}} \sqrt{\frac{\sigma^2\tau^{-2}}{\prod_{j=1,j\neq k}^{2^{m+1}} n_j} + \sum_{i=1,i\neq k}^{2^{m+1}} \frac{Z_i}{\prod_{j=1,j\neq i,j\neq k}^{2^{m+1}} n_j}}}, \qquad k = 1, 2, \cdots, 2^{m+1}. \tag{10}$$

Let $\nu = \frac{\sigma^2}{n\tau^2}$ representing the relative importance of variance versus bias. For a simple case when the model (2) with $\nu \ll \frac{n_1}{n} Z_1 + \cdots + \frac{n_{2^{m+1}}}{n} Z_{2^{m+1}}$, we have

$$\max_{g\in G_1} \Phi_{\mathbf{D}}\left(g,\xi\right) \approx \frac{\sigma^{2^{m+2}-2}\tau^2}{\left(2^{m+1}\right)^{2^{m+1}} Z_1 Z_2 \cdots Z_{2^{m+1}}} \frac{n_1 Z_1 + \cdots + n_{2^{m+1}} Z_{2^{m+1}}}{n_1 n_2 \cdots n_{2^{m+1}}}. \tag{11}$$

Consequently, we have

Corollary 1 *When ν is very small, the minimum value of (11) is attained when the design points $x_{ij} = x_{(i)}$ $\left(i = 1, \cdots, 2^{m+1}; j = 1, \cdots, n_i\right)$, where $x_{(i)}$ are the points such that $\pi\left(x_{(i)}\right) = Z_i$, and $n_1, n_2, \cdots, n_{2^{m+1}}$ are positive real numbers that satisfy*

$$\frac{n_k}{n} = \frac{\sqrt{\sum\limits_{i=1, i\neq k}^{2^{m+1}} \frac{Z_i}{\prod\limits_{j=1, j\neq i, j\neq k}^{2^{m+1}} n_j}}}{\sum\limits_{k=1}^{2^{m+1}} \sqrt{\sum\limits_{i=1, i\neq k}^{2^{m+1}} \frac{Z_i}{\prod\limits_{j=1, j\neq i, j\neq k}^{2^{m+1}} n_j}}}, \qquad k = 1, 2, \cdots, 2^{m+1}. \tag{12}$$

The algorithm for obtaining the D-minimax designs using (10) or (12) is as follows:

1. Give the starting design for n_i's: $n_1^{(0)}, n_2^{(0)}, \cdots, n_{2^{m+1}}^{(0)}$. For example, uniform design can be used as a starting design.
2. Use the recursive formula (10) or (12) depending on the value of ν to calculate $n_1^{(1)}$ by plugging in $n_1^{(0)}, n_2^{(0)}, \cdots, n_{2^{m+1}}^{(0)}$, and replace $n_1^{(0)}$ with $n_1^{(1)}$, then use the recursive formula to calculate $n_2^{(1)}$ by plugging in $n_1^{(1)}, n_2^{(0)}, \cdots, n_{2^{m+1}}^{(0)}$, and replace $n_2^{(0)}$ with $n_2^{(1)}$. Repeat this procedure to get $n_3^{(1)}, \cdots, n_{2^{m+1}-1}^{(1)}$, then replace $n_i^{(0)}$ $\left(i = 3, \cdots, 2^{m+1}-1\right)$ with $n_i^{(1)}$ $\left(i = 3, \cdots, 2^{m+1}-1\right)$. Finally, replace $n_{2^{m+1}}^{(0)}$ with $n_{2^{m+1}}^{(1)} = n - \sum\limits_{i=1}^{2^{m+1}-1} n_i^{(1)}$. This concludes the first round of our process.
3. For the kth round, we have $n_1^{(k)}, n_2^{(k)}, \cdots, n_{2^{m+1}}^{(k)}$. Then check if

$$\max_i \left(\left|\frac{n_i^{(k)} - n_i^{(k-1)}}{n}\right|\right) < \delta, \tag{13}$$

is true, where δ is a given constant (we use $\delta = \frac{1}{2^{m+1} * 1000}$ for our examples). We stop our process and output $\frac{n_1^{(k)}}{n}, \frac{n_2^{(k)}}{n}, \cdots, \frac{n_{2^{m+1}}^{(k)}}{n}$ as the final result if (13) is satisfied; otherwise we repeat steps 2 and 3.

Example 1: When $m = 2$, the approximate Haar wavelet model (2) is

$$\begin{aligned} y(x) = {} & \beta_0 + \beta_{00}\psi_{00}(x) + \beta_{10}\psi_{10}(x) + \beta_{11}\psi_{11}(x) + \beta_{20}\psi_{20}(x) \\ & + \beta_{21}\psi_{21}(x) + \beta_{22}\psi_{22}(x) + \beta_{23}\psi_{23}(x) + g(x) + \varepsilon(x), \end{aligned}$$

where $g(x) \in G_1$ and the error term $\varepsilon(x) \in \Upsilon$. Let γ be an arbitrary small positive number. If $\pi(x) = x$ and ν can be neglected, then the D-minimax design for the model is

$$\xi^{D*} \approx \begin{pmatrix} \frac{1}{8}-\gamma & \frac{2}{8}-\gamma & \frac{3}{8}-\gamma & \frac{4}{8}-\gamma & \frac{5}{8}-\gamma & \frac{6}{8}-\gamma & \frac{7}{8}-\gamma & \frac{8}{8}-\gamma \\ 0.138 & 0.134 & 0.130 & 0.126 & 0.123 & 0.119 & 0.116 & 0.113 \end{pmatrix};$$

If ν can not be neglected, suppose $\nu = 1$, then the D-minimax design for the model is

$$\xi^{D} \approx \begin{pmatrix} \frac{1}{8}-\gamma & \frac{2}{8}-\gamma & \frac{3}{8}-\gamma & \frac{4}{8}-\gamma & \frac{5}{8}-\gamma & \frac{6}{8}-\gamma & \frac{7}{8}-\gamma & \frac{8}{8}-\gamma \\ 0.129 & 0.128 & 0.127 & 0.126 & 0.124 & 0.123 & 0.122 & 0.121 \end{pmatrix}.$$

When $\pi(x) = x^2$ and ν can be neglected, then the D-minimax design for the model is

$$\xi^{D*} \approx \begin{pmatrix} \frac{1}{8}-\gamma & \frac{2}{8}-\gamma & \frac{3}{8}-\gamma & \frac{4}{8}-\gamma & \frac{5}{8}-\gamma & \frac{6}{8}-\gamma & \frac{7}{8}-\gamma & \frac{8}{8}-\gamma \\ & & & & & & & \\ 0.142 & 0.139 & 0.135 & 0.130 & 0.124 & 0.117 & 0.110 & 0.103 \end{pmatrix};$$

If ν can not be neglected, suppose $\nu = 1$, then the D-minimax design for the model is

$$\xi^{D} \approx \begin{pmatrix} \frac{1}{8}-\gamma & \frac{2}{8}-\gamma & \frac{3}{8}-\gamma & \frac{4}{8}-\gamma & \frac{5}{8}-\gamma & \frac{6}{8}-\gamma & \frac{7}{8}-\gamma & \frac{8}{8}-\gamma \\ & & & & & & & \\ 0.129 & 0.129 & 0.128 & 0.127 & 0.125 & 0.123 & 0.121 & 0.118 \end{pmatrix}.$$

When $\pi(x) = \exp(-x)$ and ν can be neglected, then the D-minimax design for the model is

$$\xi^{D*} \approx \begin{pmatrix} \frac{0}{8} & \frac{1}{8} & \frac{2}{8} & \frac{3}{8} & \frac{4}{8} & \frac{5}{8} & \frac{6}{8} & \frac{7}{8} \\ & & & & & & & \\ 0.118 & 0.120 & 0.122 & 0.124 & 0.126 & 0.129 & 0.130 & 0.131 \end{pmatrix};$$

If ν can not be neglected, suppose $\nu = 1$, then the D-minimax design is

$$\xi^{D} \approx \begin{pmatrix} \frac{0}{8} & \frac{1}{8} & \frac{2}{8} & \frac{3}{8} & \frac{4}{8} & \frac{5}{8} & \frac{6}{8} & \frac{7}{8} \\ & & & & & & & \\ 0.122 & 0.123 & 0.124 & 0.125 & 0.126 & 0.126 & 0.127 & 0.127 \end{pmatrix}.$$

We note that the convergences in obtaining D-minimax designs have been reached very quickly in all situations we have tested for various efficiency functions, and various values of ■, even with a larger m. A list of convergence times is given in Table 1.

3.2 Equivalency of Q-minimax and A-minimax designs for the Approximate Haar wavelet Models over G_1

Q-minimax designs are sensible when we are interested in prediction. In this case, the problem becomes how to find the best estimation of the response function; the corresponding loss function is the average mean squared prediction error:

$$\Phi_{\mathbf{Q}}(g,\xi) = \int_{x \in S} \mathbf{z}(\mathbf{x})^T \mathbf{M}(g, \boldsymbol{\xi}) \mathbf{z}(\mathbf{x})\, dx.$$

With (5), (6), (7) and (8) in the previous subsection, we have

$$\Phi_{\mathbf{Q}}(g,\xi) = 2^{-2(m+1)} \begin{bmatrix} \sigma^2 \int_0^1 \left[\mathbf{z}(x)^T \mathbf{X}_{\mathbf{H}}^T \mathbf{L}^{-1} \mathbf{X}_{\mathbf{H}} \mathbf{z}(x)\right] dx \\ + \int_0^1 \left[\mathbf{z}(x)^T \mathbf{X}_{\mathbf{H}}^T \mathbf{g}_B \mathbf{g}_B^T \mathbf{X}_{\mathbf{H}} \mathbf{z}(x)\right] dx \end{bmatrix}, \tag{14}$$

where

$$\mathbf{z}(x) = \begin{pmatrix} 1 \\ \psi_{00}(x) \\ \vdots \\ \psi_{m,2^m-1}(x) \end{pmatrix}.$$

We note that

$$\int_0^1 \mathbf{z}(x)\, g(x)\, dx = \mathbf{0}, \ \int_0^1 \mathbf{z}(x)\, \mathbf{z}^T(x)\, dx = \mathbf{I}.$$

Let $Q_1 = \int_0^1 \left[\mathbf{z}(x)^T \mathbf{X}_{\mathbf{H}}^T \sigma^2 \mathbf{L}^{-1} \mathbf{X}_{\mathbf{H}} \mathbf{z}(x)\right] dx$, and $Q_2 = \int_0^1 \left[\mathbf{z}(x)^T \mathbf{X}_{\mathbf{H}}^T \mathbf{g}_B \mathbf{g}_B^T \mathbf{X}_{\mathbf{H}} \mathbf{z}(x)\right] dx$. We have

$$\begin{aligned}
\max_{g\in G_1} \Phi_{\mathbf{Q}}(g,\xi) &= 2^{-2(m+1)} \left[Q_1 + \max_{g\in G_1} Q_2\right] \\
&= 2^{-2(m+1)} \left[Q_1 + \max_{g\in G_1} \mathbf{g}_B^T \left(\int_0^1 \mathbf{X}_{\mathbf{H}} \mathbf{z}(x) \mathbf{z}(x)^T \mathbf{X}_{\mathbf{H}}^T dx\right) \mathbf{g}_B\right] \\
&= 2^{-2(m+1)} \left[Q_1 + 2^{m+1} \tau^2 \lambda_{\max} \left(\int_0^1 \mathbf{X}_{\mathbf{H}} \mathbf{z}(x) \mathbf{z}(x)^T \mathbf{X}_{\mathbf{H}}^T dx\right)\right],
\end{aligned}$$

with

$$\mathbf{X}_{\mathbf{H}} \mathbf{z}(x) = \begin{pmatrix} 1 + \psi_{00}(u_1)\psi_{00}(x) + \cdots + \psi_{m,2^m-1}(u_1)\psi_{m,2^m-1}(x) \\ 1 + \psi_{00}(u_2)\psi_{00}(x) + \cdots + \psi_{m,2^m-1}(u_2)\psi_{m,2^m-1}(x) \\ \vdots \\ 1 + \psi_{00}(u_{2^{m+1}})\psi_{00}(x) + \cdots + \psi_{m,2^m-1}(u_{2^{m+1}})\psi_{m,2^m-1}(x) \end{pmatrix}.$$

Suppose $s, w = 1, 2, \cdots, 2^{m+1}$, we have $\mathbf{X}_{\mathbf{H}} \mathbf{z}(x) \mathbf{z}(x)^T \mathbf{X}_{\mathbf{H}}^T$ being a $2^{m+1} \times 2^{m+1}$ matrix with elements of

$$a_{s,w} = \left\{\left(1 + \sum_{j=0}^{m} \sum_{k=0}^{2^j-1} \psi_{jk}(u_s)\psi_{jk}(x)\right)\left(1 + \sum_{j=0}^{m} \sum_{k=0}^{2^j-1} \psi_{jk}(u_w)\psi_{jk}(x)\right)\right\}_{s,w},$$

and

$$\begin{aligned}
&\lambda_{\max}\left(\int_0^1 \mathbf{X}_{\mathbf{H}} \mathbf{z}(x) \mathbf{z}(x)^T \mathbf{X}_{\mathbf{H}}^T dx\right) \\
&= \lambda_{\max}\left(\begin{bmatrix} 2^{m+1} & 0 & 0 & 0 \\ 0 & 2^{m+1} & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & 2^{m+1} \end{bmatrix}\right) = 2^{m+1},
\end{aligned}$$

where $\int_0^1 \mathbf{X}_{\mathbf{H}} \mathbf{z}(x) \mathbf{z}(x)^T \mathbf{X}_{\mathbf{H}}^T dx$ is the corresponding $2^{m+1} \times 2^{m+1}$ matrix with elements of $\int_0^1 a_{s,w} dx$. Therefore,

$$\max_{g\in G_1} \Phi_{\mathbf{Q}}(g,\xi) = 2^{-2(m+1)} \int_0^1 \left[\mathbf{z}(x)^T \mathbf{X}_{\mathbf{H}}^T \sigma^2 \mathbf{L}^{-1} \mathbf{X}_{\mathbf{H}} \mathbf{z}(x)\right] dx + \tau^2.$$

Now, our goal of minimizing $\max_{g\in G_1} \Phi_{\mathbf{Q}}(g,\xi)$ becomes to minimize $\int_0^1 \left[\mathbf{z}(x)^T \mathbf{X}_{\mathbf{H}}^T \mathbf{L}^{-1} \mathbf{X}_{\mathbf{H}} \mathbf{z}(x)\right] dx$. The Q-minimax designs can be constructed by the following theorem:

Theorem 2 *The minimum of mean squared prediction error,* $\max_{g\in G_1} \Phi_{\mathbf{Q}}(g,\xi)$*, attains its minimum value when the design points* $x_{ij} = x_{(i)}$ $\left(i = 1, \cdots, 2^{m+1}; j = 1, \cdots, n_i\right)$*, where* $x_{(i)}$ *are the points such that* $\pi\left(x_{(i)}\right) = Z_i$*, and* $n_1, n_2, \cdots, n_{2^{m+1}}$ *are positive real numbers which satisfy*

$$\frac{n_k}{n} = \frac{\sqrt{\frac{1}{Z_k}}}{\sum_{i=1}^{2^{m+1}} \left(\sqrt{\frac{1}{Z_i}}\right)}, \qquad k = 1, 2, \cdots, 2^{m+1}. \tag{15}$$

Comparing to the A-minimax design found in Tian and Herzberg (2006B), Q-minimax designs are equivalent to A-optimal designs for the model (2) over G_1.

4 Robust Designs for the Approximate Haar wavelet Models over G_2

In this section, we consider a broader contamination class G_2 below

$$G_2 = \left\{ g(x) \mid \int_0^1 g^2(x)\,dx \le \tau^2, \ \int_0^1 g(x)\,dx = 0 \text{ and } \int_0^1 \psi_{jk}(x)\,g(x)\,dx = 0 \right\}. \tag{16}$$

This class was first adopted first by Huber (1975), and the virtues of this class have further been indicated in Wiens (1992). We now discuss the D-, A-, and Q-optimal robust designs under G_2.

Oyet and Wiens (2000) consider the approximate wavelet model (2) with $g(x) \in G_2$ and homogenous errors satisfy:

$$E(\varepsilon) = 0,\ Var(\varepsilon) = \sigma^2 \text{ and } Cov\left[\varepsilon(x), \varepsilon(x')\right] = 0 \text{ for } x \neq x'.$$

They indicate that, under G_2, only absolute continuous designs $\xi(x)$ have finite maximum loss; and also conclude (i) that the A-, D-, Q- minimax robust designs are equivalent, and (ii) that uniform design is the minimax robust design for all these three loss functions. In this section, we extend the results of Oyet and Wiens (2000) for the case of homoscedasticity to the case of heteroscedasticity. We consider the model (2) with $g(x) \in G_2,$ and heterogeneous errors $\varepsilon(x) \in \Upsilon$ instead.

We make use of the following matrices and vectors:

$$\mathbf{B} = \mathbf{B}(\xi) = \frac{\mathbf{X}_\mathbf{H}^T \mathbf{X}_\mathbf{H}}{n} = \frac{1}{n}\sum_{i=1}^n \mathbf{z}(x_i)\,\mathbf{z}^T(x_i) = \int_S \mathbf{z}(x)\,\mathbf{z}^T(x)\,d\xi(x),$$

$$\mathbf{b} = \mathbf{b}(g,\xi) = \frac{\mathbf{X}_\mathbf{H}^T \mathbf{g}(x)}{n} = \frac{1}{n}\sum_{i=1}^n \mathbf{z}(x_i)\,g(x_i) = \int_S \mathbf{z}(x)\,g(x)\,d\xi(x),$$

$$\mathbf{D} = \int_S \mathbf{z}(x)\,\mathbf{z}^T(x)\,\pi^{-1}(x)\,d\xi(x).$$

In this case, the loss functions for the approximate Haar wavelet models become

(1). Q-optimal:

$$Qloss = \mathbf{b}^T\mathbf{B}^{-2}\mathbf{b} + \frac{\sigma^2}{n} tr\mathbf{B}^{-1}\mathbf{D}\mathbf{B}^{-1} + \int_S g^2(x)\,dx.$$

(2). A-optimal:

$$\begin{aligned} Aloss &= trM(g,\xi) \\ &= \mathbf{b}^T\mathbf{B}^{-2}\mathbf{b} + \frac{\sigma^2}{n} tr\mathbf{B}^{-1}\mathbf{D}\mathbf{B}^{-1}. \end{aligned}$$

(3). D-optimal:

$$\begin{aligned} Dloss &= \left|M(g,\xi)\right| \\ &= \left|\mathbf{B}^{-1}\mathbf{b}\mathbf{b}^T\mathbf{B}^{-1} + \frac{\sigma^2}{n}\mathbf{B}^{-1}\mathbf{D}\mathbf{B}^{-1}\right| \\ &= \left|\mathbf{B}^{-1}\right| \cdot \left|\mathbf{b}\mathbf{b}^T + \frac{\sigma^2}{n}\mathbf{D}\right| \cdot \left|\mathbf{B}^{-1}\right| \\ &= \left(\frac{\sigma^2}{n}\right)^{2^{m+1}} \left[\frac{\left[1 + \left(\frac{\sigma^2}{n}\right)^{-1}\mathbf{b}^T\mathbf{D}^{-1}\mathbf{b}\right]\left|\mathbf{D}\right|}{\left|\mathbf{B}\right|^2}\right]. \end{aligned}$$

Let $m(x) = \xi'(x)$, $\mathbf{C} = \int_\mathbf{S} \mathbf{z}(x)\,\mathbf{z}^T(x)\,m^2(x)\,\mathbf{d}x$, $\mathbf{K} = \int_\mathbf{S} \mathbf{z}(x)\,\mathbf{z}^T(x)\,m^2(x)\,\mathbf{d}x - \mathbf{B}^2 = \mathbf{C} - \mathbf{B}^2$. The maxima of the loss functions above are stated in the following theorem:

Theorem 3 *For model (2), the maxima of Qloss, Aloss and Dloss over* G_2 *are*

$$\max_{g\in G_2} Qloss = \tau^2 \left(\lambda_{\max}\left(\mathbf{KB}^{-2}\right) + \nu \cdot tr\mathbf{B}^{-1}\mathbf{DB}^{-1} + 1\right), \tag{17}$$

$$\max_{g\in G_2} Aloss = \tau^2 \left(\lambda_{\max}\left(\mathbf{KB}^{-2}\right) + \nu \cdot tr\mathbf{B}^{-1}\mathbf{DB}^{-1}\right), \tag{18}$$

and

$$\max_{g\in G_2} Dloss = \tau^2 \left(\frac{\sigma^2}{n}\right)^{2^{m+1}-1} \left(\frac{\left[\nu + \lambda_{\max}\left(\mathbf{KD}^{-1}\right)\right]|\mathbf{D}|}{|\mathbf{B}|^2}\right). \tag{19}$$

It can be shown that for heteroscedasticity case, the D-minimax design is not equivalent to A- or Q-minimax design any more. In addition, neither A- nor Q- minimax design is a uniform design. The following theorem provides the D-minimax, A-minimax, and Q-minimax designs for heteroscedasticity case.

Theorem 4 *(1) D-minimax designs over* G_2 *are uniform designs, that is* $m(x) = 1$*; (2) The densities of A-minimax and Q-minimax designs have the same form as:*

$$m(x) = \left(\frac{\mathbf{z}^T(x)\,\mathbf{Mz}(x) - \pi^{-1}(x)\,\mathbf{z}^T(x)\,\mathbf{Nz}(x) - t}{\mathbf{z}^T(x)\,\mathbf{Uz}(x)}\right)^+, \tag{20}$$

where three symmetric constant matrices $\mathbf{M}$*,* $\mathbf{N}\,(\geq 0)$*,* $\mathbf{U}\,(> 0)$ *and a scalar* t *can be determined by minimizing (18).*

Because uniform design is a D-minimax design and $\mathbf{z}(x)$ does not change when $x_i \in \chi_i = \left[\frac{i-1}{2^{m+1}}, \frac{i}{2^{m+1}}\right)$, we simply choose $x_i = \frac{i-1}{2^{m+1}} + \frac{1}{2^{m+2}}$, $i = 1, 2, \cdots, 2^{m+1}$. In this case, we give the following examples.

Example 3: We have

$$\xi^D = \begin{pmatrix} \frac{1}{8} & \frac{3}{8} & \frac{5}{8} & \frac{7}{8} \\ \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \end{pmatrix}$$

for $m = 1$, while

$$\xi^D = \begin{pmatrix} \frac{1}{16} & \frac{3}{16} & \frac{5}{16} & \frac{7}{16} & \frac{9}{16} & \frac{11}{16} & \frac{13}{16} & \frac{15}{16} \\ \\ \frac{1}{8} & \frac{1}{8} & \frac{1}{8} & \frac{1}{8} & \frac{1}{8} & \frac{1}{8} & \frac{1}{8} & \frac{1}{8} \end{pmatrix},$$

when $m = 2$.

We note that for multiwavelet approximation or regression approximated by other wavelets with higher vanishing moments, the analytic form of (20) stays the same except with different $\mathbf{z}(x)$.

Numerical solutions on A-minimax and Q-minimax designs will be discussed in the next section.

5 Examples, Discretization, and Simulation

The general theoretical solution on A- and Q-minimax designs found for the approximate Haar wavelet models over G_2 in Theorem 5 has reduced the problem to one of numerical minimization. First, we will discuss the numerical solutions on A- and Q-minimax designs by using Splus; then we will conduct a simulation for A-minimax designs over G_2.

5.1 Numerical Solution on A-, Q-minimax Designs over G_2

In Section 4, we have proven that A- and Q-minimax designs have the same form as in (20). Now we will use the simplest case of $m = 0$ to show the resulting numerical solutions on A- and Q-minimax designs for various over G_2. It is obvious that $m(x)$ can be simplified as:

$$m(x) = \begin{cases} \left(a_1 - a_2\pi^{-1}(x)\right)^+, & x \in \left[0, \frac{1}{2}\right); \\ \left(b_1 - b_2\pi^{-1}(x)\right)^+, & x \in \left[\frac{1}{2}, 1\right). \end{cases} \tag{21}$$

We let

$$\begin{gathered} d_{11} = \int_0^{\frac{1}{2}} m(x)\,dx,\ d_{12} = \int_{\frac{1}{2}}^1 m(x)\,dx, \\ d_{21} = \int_0^{\frac{1}{2}} m^2(x)\,dx,\ d_{22} = \int_{\frac{1}{2}}^1 m^2(x)\,dx, \\ ld_1 = \int_0^{\frac{1}{2}} \pi^{-1}(x)\,m(x)\,dx,\ ld_2 = \int_{\frac{1}{2}}^1 \pi^{-1}(x)\,m(x)\,dx, \end{gathered} \tag{22}$$

where $d_{11} + d_{12} = 1$. Then we can obtain the following maximum A-loss function:

$$\max_{g \in G_2} Aloss = \begin{cases} \tau^2\left(\frac{d_{22}d_{11}^2 + d_{21}d_{12}^2 - 4d_{11}^2d_{12}^2}{4d_{11}^2d_{12}^2} + \frac{d_{22}d_{11}^2 - d_{21}d_{12}^2}{4d_{11}^2d_{12}^2} + \nu\frac{ld_2d_{11}^2 + ld_1d_{12}^2}{2d_{11}^2d_{12}^2}\right), & d_{22}d_{11}^2 - d_{21}d_{12}^2 \geq 0 \\ \tau^2\left(\frac{d_{22}d_{11}^2 + d_{21}d_{12}^2 - 4d_{11}^2d_{12}^2}{4d_{11}^2d_{12}^2} - \frac{d_{22}d_{11}^2 - d_{21}d_{12}^2}{4d_{11}^2d_{12}^2} + \nu\frac{ld_2d_{11}^2 + ld_1d_{12}^2}{2d_{11}^2d_{12}^2}\right), & d_{22}d_{11}^2 - d_{21}d_{12}^2 < 0 \end{cases}. \tag{23}$$

Plugging (21) into (23), we obtain the numerical solution of the coefficients in (20) for a few examples as follows.

We take $\nu = 0.1$, 1, and 5; also take $\pi(x) = x$, $\frac{1}{x}$, e^x, and e^{-x} in the examples. The design density $m(x)$ is in form of (21). For various ν and $\pi(x)$, the coefficients a_1, a_2, b_1, and b_2 in (21) are shown in Tables 2 and 3. The graphs of these design densities are shown in Figure 1, 2, 3, and 4.

The resulting design densities suggest that an increasing efficiency function induces a piecewise increasing design density (such as those in Figure 1 and 3) while a decreasing efficiency function results in a piecewise decreasing design density (such as those in Figure 2 and 4). This means that relatively more test subjects should be allocated in those areas that have larger values of efficiency functions. In addition, decreasing ■, meaning less effort on minimizing variances than that on reducing biases, causes more uniform designs.

The relative efficiency of the design obtained relative to uniform design is discussed next. We denote the minimum of Aloss function as $MinAloss$, and denote the loss function, when uniform design is applied, as $Uloss$. We denote the relative efficiency as eff, which is defined by $\frac{Uloss}{MinAloss}$. The $MinAloss$, $Uloss$, and eff are provided in Tables 4 and 5 for various $\pi(x)$.

5.2 Discretization

In order to implement the designs with densities, discretization has to be taken place. We use the method of matching quantiles and find those design support points x_i such that

$$\int_0^{x_i} m(x)\,dx = \frac{i - 0.5}{10}, i = 1, 2, \cdots, n.$$

In this section we take the sample size of $n = 10$ as an example. For various $\pi(x)$, the resulting design points after discretization are listed in Table 6. Also see Figure 5, 6, 7, and 8 for their plots.

5.3 Simulation

Finally, we do a simulation by using the true model

$$y(x) = 1 - 0.5\psi_{00}(x) + g(x) + \varepsilon(x),$$

where $g(x) \in G_2$, $\varepsilon(x) \in \Upsilon$. Suppose $\pi(x) = \frac{1}{x}$, $\tau^2 = 0.1$, $\sigma^2 = 5$, $n = 10$, and $g(x)$ is

$$g(x) = \begin{cases} \frac{\sqrt{5}}{2}, & x \in [0.1090, 0.1190], \\ \frac{\sqrt{5}}{2}, & x \in [0.2190, 0.2290], \\ \frac{\sqrt{5}}{2}, & x \in [0.5090, 0.5190], \\ \frac{\sqrt{5}}{2} & x \in [0.5400, 0.5500] \\ \frac{\sqrt{5}}{2} & x \in [0.5800, 0.5900] \\ \frac{\sqrt{5}}{2} & x \in [0.6190, 0.6290] \\ \frac{\sqrt{5}}{2} & x \in [0.6700, 0.6800] \\ \frac{\sqrt{5}}{2} & x \in [0.8090, 0.8190] \\ 0 & \text{otherwise.} \end{cases}$$

When $\nu = 5$, the A-minimax design found in Table 6 is

$$\begin{pmatrix} 0.1175 & 0.2290 & 0.3550 & 0.5090 & 0.5430 & 0.5800 & 0.6220 & 0.6705 & 0.7295 & 0.8160 \end{pmatrix}.$$

The uniform design is:

$$\begin{pmatrix} 0 & 0.1 & 0.2 & 0.3 & 0.4 & 0.5 & 0.6 & 0.7 & 0.8 & 0.9 \end{pmatrix}.$$

We perform this simulation 200 times. The values of $Uloss$ and $MinAloss$ are shown in Appendix I. We also calculate $AMinAloss$, which is the average of MinAloss over these 200 runs, and $AUloss$, which is the average of Uloss over these 200 runs. The efficiency eff is defined as $\frac{AUloss}{AMinloss}$, and the simulation results are as follows:

ν	$AMinAloss$	$AUloss$	eff
5	131.63	156.55	1.189319

This simulation study confirms that our robust design is more efficient than uniform design when the model is possibly contaminated.

6 Concluding Remarks

This paper presents the methods of constructing optimally robust designs for wavelet regression in the cases of heteroscedasticity, taking the uncertainty in assumed regression function into account. Some of the resulting designs require possible extensive numerical work prior to implementation. Nevertheless, the results of this paper provide valuable guidelines in designing an experiment for wavelet regression problems.

Acknowledgements

The research of both authors is supported by the Natural Sciences and Engineering Research Council of Canada. The authors also thank the anonymous referees for their helpful comments on the previous version of this article.

Appendix: Derivations

Proof of Theorem 1: Let

$$\Phi\left(n_1, n_2, \cdots, n_{2^{m+1}}, t\right)=\frac{\sigma^2 \tau^{-2}+n_1 Z_1+\cdots+n_{2^{m+1}} Z_{2^{m+1}}}{n_1 n_2 \cdots n_{2^{m+1}}}+t\left(n_1+n_2+\cdots+n_{2^{m+1}}-n\right).$$

We need to solve the equations below

$$\begin{cases}\frac{\partial \Phi\left(n_1, n_2, \cdots, n_{2^{m+1}}, t\right)}{\partial n_1}=0, \\ \frac{\partial \Phi\left(n_1, n_2, \cdots, n_{2^{m+1}}, t\right)}{\partial n_2}=0, \\ \vdots \\ \frac{\partial \Phi\left(n_1, n_2, \cdots, n_{2^{m+1}}, t\right)}{\partial t}=0.\end{cases},$$

$\Phi\left(n_1, n_2, \cdots, n_{2^{m+1}}, t\right)$ can be written as

$$\frac{\sigma^2 \tau^{-2}}{\prod_{j=1}^{2^{m+1}} n_j}+\sum_{i=1}^{2^{m+1}} \frac{Z_i}{\prod_{j=1, j \neq i}^{2^{m+1}} n_j}+t\left(\sum_{i=1}^{2^{m+1}} n_i-n\right),$$

and the partial derivatives of $\Phi\left(n_1, n_2, \cdots, n_{2^{m+1}}, t\right)$ are

$$\begin{cases}\frac{\partial \Phi^{\prime}\left(n_1, n_2, \cdots, n_{2^{m+1}}, \gamma\right)}{\partial n_k}=-\frac{1}{n_k^2}\left(\frac{\sigma^2 \tau^{-2}}{\prod_{j=1, j \neq k}^{2^{m+1}} n_j}+\sum_{i=1, i \neq k}^{2^{m+1}} \frac{Z_i}{\prod_{j=1, j \neq i, j \neq k}^{2^{m+1}} n_j}\right)+t, \ k=1,2, \cdots, 2^{m+1}; \\ \frac{\partial \Phi^{\prime}\left(n_1, n_2, \cdots, n_{2^{m+1}}, \gamma\right)}{\partial t}=\sum_{i=1}^{2^{m+1}} n_i-n.\end{cases}$$

Set these equations to zero, then (10) is attained. □

Proof of Corollary 1: is very similar to that of Theorem 1, so we omit it.

Proof of Theorem 2: We minimize $\int_0^1\left[\mathbf{z}(x)^T \mathbf{X}_{\mathbf{H}}^T \mathbf{L}^{-1} \mathbf{X}_{\mathbf{H}} \mathbf{z}(x)\right] dx$ over design support points (x_i) and allocations (n_i). We have

$$\begin{aligned}
& \min_{\xi} \int_0^1\left[\mathbf{z}(x)^T \mathbf{X}_{\mathbf{H}}^T \mathbf{L}^{-1} \mathbf{X}_{\mathbf{H}} \mathbf{z}(x)\right] dx \\
= & \min_{\xi} \int_0^1\left[\sum_{i=1}^{2^{m+1}}\left(\frac{\left[1+\sum_{j=0}^m \sum_{k=0}^{2^j-1} \psi_{jk}\left(u_i\right) \psi_{jk}(x)\right]^2}{\sum_{j=1}^{n_i} \pi\left(x_{ij}\right)}\right)\right] dx \\
= & \min_{\xi}\left[\sum_{i=1}^{2^{m+1}}\left(\frac{\int_0^1\left[1+\sum_{j=0}^m \sum_{k=0}^{2^j-1} \psi_{jk}\left(u_i\right) \psi_{jk}(x)\right]^2 dx}{\sum_{j=1}^{n_i} \pi\left(x_{ij}\right)}\right)\right].
\end{aligned}$$

In order to minimize over x_i, all $\pi(x_{ij})$ must satisfy that $\pi(x_{ij}) = Z_i = \pi\left(x_{(i)}\right)$, where $\chi_i = \left[\frac{i-1}{2^{m+1}}, \frac{i}{2^{m+1}}\right)$, $i = 1, \cdots, 2^{m+1}$, and then

$$\min_{\xi} \int_0^1 \left[\mathbf{z}(x)^T \mathbf{X}_{\mathbf{H}}^T \mathbf{L}^{-1} \mathbf{X}_{\mathbf{H}} \mathbf{z}(x)\right] dx$$

$$= \min_{n_i} \left[\sum_{i=1}^{2^{m+1}} \left(\frac{\int_0^1 \left[1 + \sum_{j=0}^{m} \sum_{k=0}^{2^j-1} \psi_{jk}(u_i)\psi_{jk}(x)\right]^2 dx}{n_i Z_i} \right)\right].$$

Because

$$\int_0^1 \left[1 + \sum_{j=0}^{m} \sum_{k=0}^{2^j-1} \psi_{jk}(u_i)\psi_{jk}(x)\right]^2 dx = \left(2^{m+1}\right)^2 \frac{1}{2^{m+1}} = 2^{m+1},$$

the minimum over n_i is

$$\min_{n_i} \int_0^1 \left[\mathbf{z}(x)^T \mathbf{X}_{\mathbf{H}}^T \mathbf{L}^{-1} \mathbf{X}_{\mathbf{H}} \mathbf{z}(x)\right] dx = \min_{n_i} \left[\sum_{i=1}^{2^{m+1}} \left(\frac{2^{m+1}}{n_i Z_i}\right)\right],$$

subject to $n_1 + n_2 + \cdots + n_{2^{m+1}} = n$. Then, this theorem follows.

Proof of Theorem 3: We omit the proofs for (17) and (18) because they are very similar to those in Oyet and Wiens (2000). In order to get $\max\limits_{g \in G_2} Dloss$, which is the maximum of Dloss, we use the method of Lagrange multipliers to maximize $\mathbf{b}^T\mathbf{D}^{-1}\mathbf{b}$, where $\mathbf{b} = \int_0^1 \mathbf{f}(x) g(x) m(x)\, dx$, $\mathbf{D} = \int_0^1 \mathbf{f}(x) \mathbf{f}^T(x) \pi^{-1}(x) m(x)\, dx$, and

$$\int_0^1 \mathbf{f}(x) g(x)\, dx = \mathbf{0}, \quad \int_0^1 g^2(x)\, dx \le \tau^2.$$

If $\int_0^1 g_k^2(x)\, dx = k^2 < \tau^2$, let $g_\tau(x) = \frac{\tau}{k} g_k(x)$, meaning we have $\int_0^1 g_\tau^2(x)\, dx = \tau^2$ and $\mathbf{b}^T(g_k)\mathbf{D}^{-1}\mathbf{b}(g_k) = \frac{k^2}{\tau^2}\mathbf{b}^T(g_\tau)\mathbf{D}^{-1}\mathbf{b}(g_\tau) < \mathbf{b}^T(g_\tau)\mathbf{D}^{-1}\mathbf{b}(g_\tau)$. Therefore, the maximum loss can be searched over $g(x) \in G_2^*$ where

$$G_2^* = \left\{ g(x) \mid \int_0^1 g^2(x)\, dx = \tau^2, \quad \int_0^1 g(x)\, dx = 0 \text{ and } \int_0^1 \psi_{jk}(x) g(x)\, dx = 0 \right\}.$$

Let

$$F = \mathbf{b}^T\mathbf{D}^{-1}\mathbf{b} - \gamma_1 \left(\int_0^1 g^2(x)\, dx - \tau^2\right) - 2\left(\int_0^1 \mathbf{f}^T(x) g(x)\, dx\right) \boldsymbol{\gamma}_2,$$

then let

$$\frac{\partial F}{\partial g} = 0,$$

we have

$$\begin{cases} g(x) = \frac{\mathbf{b}^T\mathbf{D}^{-1}\mathbf{f}(x)m(x) - \mathbf{f}^T(x)\boldsymbol{\gamma}_2}{\gamma_1}, \\ \int_0^1 \mathbf{f}(x) g(x)\, dx = \mathbf{0}, \\ \int_0^1 g^2(x)\, dx = \tau^2. \end{cases} \tag{A.1}$$

By solving the equations (A.1) we obtain:

$$\begin{cases} \int_0^1 \mathbf{f}(x) \left[\frac{\mathbf{b}^T\mathbf{D}^{-1}\mathbf{f}(x)m(x) - \mathbf{f}^T(x)\boldsymbol{\gamma}_2}{\gamma_1}\right] dx = \mathbf{0}, \\ \frac{\int_0^1 \left[\mathbf{b}^T\mathbf{D}^{-1}\mathbf{f}(x)m(x) - \mathbf{f}^T(x)\boldsymbol{\gamma}_2\right]^2 dx}{\gamma_1^2} = \tau^2. \end{cases}$$

From the first equation in (**??**) we get

$$\frac{1}{\gamma_1}\left[\int_0^1 \mathbf{b}^T\mathbf{D}^{-1}\mathbf{f}(x)\,\mathbf{f}^T(x)\,m(x)\,dx-\int_0^1 \boldsymbol{\gamma}_2^T\mathbf{f}(x)\,\mathbf{f}^T(x)\,dx\right]=\mathbf{0}^T.$$

Namely

$$\boldsymbol{\gamma}_2=\mathbf{B}\mathbf{D}^{-\mathbf{1}}\mathbf{b}. \tag{A.2}$$

Plugging (A.2) into the first equation in (A.1), then we have

$$\begin{aligned} g_0(x) &= \frac{\mathbf{b}^T\mathbf{D}^{-1}\mathbf{f}(x)\,m(x)-\mathbf{b}^T\mathbf{D}^{-\mathbf{1}}\mathbf{B}\mathbf{f}(x)}{\gamma_1} \\ &= \frac{\mathbf{f}^T(x)\,(m(x)\,\mathbf{I}-\mathbf{B})\,\mathbf{D}^{-1}\mathbf{b}}{\gamma_1}=\mathbf{f}^T(x)\,[m(x)\,\mathbf{I}-\mathbf{B}]\,\mathbf{D}^{-1}\mathbf{c}, \end{aligned} \tag{A.3}$$

where $\mathbf{c}=\frac{\mathbf{b}}{\gamma_1}$. So

$$\begin{aligned} \mathbf{b} &= \int_0^1 \mathbf{f}(x)\,g(x)\,m(x)\,dx \\ &= \int_0^1 \mathbf{f}(x)\,\mathbf{f}^T(x)\,m(x)\,[m(x)\,\mathbf{I}-\mathbf{B}]\,\mathbf{D}^{-1}\mathbf{c}dx \\ &= \mathbf{K}\mathbf{D}^{-1}\mathbf{c}, \end{aligned}$$

and

$$\mathbf{b}^T\mathbf{D}^{-\mathbf{1}}\mathbf{b}=\mathbf{c}^T\mathbf{D}^{-\mathbf{1}}\mathbf{K}\mathbf{D}^{-\mathbf{1}}\mathbf{K}\mathbf{D}^{-\mathbf{1}}\mathbf{c}. \tag{A.4}$$

From the second equation in (**??**) we get

$$\int_0^1 \left\{\mathbf{c}^T\mathbf{D}^{-1}\,[m(x)\,\mathbf{I}-\mathbf{B}]\,\mathbf{f}(x)\right\}^2 dx=\tau^2.$$

Let

$$\begin{aligned} \mathbf{H} &= \int_0^1 \left[\mathbf{D}^{-1}\{m(x)\,\mathbf{I}-\mathbf{B}\}\,\mathbf{f}(x)\right]\left[\mathbf{f}^T(x)\,\{m(x)\,\mathbf{I}-\mathbf{B}\}\,\mathbf{D}^{-1}\right]dx \\ &= \mathbf{D}^{-1}\mathbf{C}\mathbf{D}^{-1}-\mathbf{D}^{-1}\mathbf{B}^2\mathbf{D}^{-1} \\ &= \mathbf{D}^{-1}\mathbf{K}\mathbf{D}^{-1}, \end{aligned}$$

where $C=\int_{\mathbf{S}}\mathbf{f}(x)\,\mathbf{f}^T(x)\,m^2(x)\,\mathbf{d}x$ and

$$\mathbf{a}=\frac{\mathbf{H}^{\frac{1}{2}}\mathbf{c}}{\tau}.$$

Then we have

$$\mathbf{c}=\tau\mathbf{H}^{-\frac{1}{2}}\mathbf{a}. \tag{A.5}$$

Plugging (A.5) into (A.3) we can get

$$g_0(x)=\tau\mathbf{f}^T(x)\,[m(x)\,\mathbf{I}-\mathbf{B}]\,\mathbf{D}^{-1}\mathbf{H}^{-\frac{1}{2}}\mathbf{a}, \tag{A.6}$$

for some $\mathbf{a}$ satisfying $\|\mathbf{a}\|^2=1$. We write (A.6) as

$$g_0(x)=\mathbf{u}^T(x)\,\mathbf{a},$$

where

$$\mathbf{u}^T(x)=\tau\mathbf{f}^T(x)\,(m(x)\,\mathbf{I}-\mathbf{B})\,\mathbf{D}^{-1}\mathbf{H}^{-\frac{1}{2}}.$$

And plugging (A.5) into (A.4) we can get

$$\mathbf{b}^T\mathbf{D}^{-1}\mathbf{b} = \tau^2\mathbf{a}^T\mathbf{H}^{-\frac{1}{2}}\mathbf{D}^{-1}\mathbf{K}\mathbf{D}^{-1}\mathbf{K}\mathbf{D}^{-1}\mathbf{H}^{-\frac{1}{2}}\mathbf{a}. \tag{A.7}$$

Now, our problem becomes to maximize (A.7) subject to $||\mathbf{a}||^2 = 1$. Because

$$\begin{aligned}\max_{||\mathbf{a}||^2=1} \tau^2\mathbf{a}^T\mathbf{H}^{-\frac{1}{2}}\mathbf{D}^{-1}\mathbf{K}\mathbf{D}^{-1}\mathbf{K}\mathbf{D}^{-1}\mathbf{H}^{-\frac{1}{2}}\mathbf{a} &= \tau^2\lambda_{\max}\left(\mathbf{H}^{-\frac{1}{2}}\mathbf{D}^{-1}\mathbf{K}\mathbf{D}^{-1}\mathbf{K}\mathbf{D}^{-1}\mathbf{H}^{-\frac{1}{2}}\right)\\ &= \tau^2\lambda_{\max}\left(\mathbf{K}\mathbf{D}^{-1}\right),\end{aligned}$$

and $\mathbf{b}^T\mathbf{D}^{-1}\mathbf{b}$ is the only term includes $g(x)$, the maximum of Dloss is (19).□

Proof of Theorem 4: (1) For D-minimax we have the non-negative definite matrix $\mathbf{K}\mathbf{D}^{-1}$, which means $\min \lambda_{\max}\left(\mathbf{K}\mathbf{D}^{-1}\right) = 0$. We obtain that the uniform design ξ^* is one of the situations where $\min \lambda_{\max}\left(\mathbf{K}\mathbf{D}^{-1}\right) = 0$. Also, we know that $\max|\mathbf{B}| = \max\left|\int_0^1 \mathbf{f}(x)\mathbf{f}^T(x)m(x)\,dx\right| = |\mathbf{I}| = 1$. Therefore, the uniform design is a D-minimax design and

$$\min_{m(x)}\max_{g\in G_2} Dloss = \left(\frac{\sigma^2}{n}\right)^{2^{m+1}}.$$

(2) For A-minimax and Q-minimax, first we can easily know from (17) and (18) that A-minimax designs and Q-minimax designs are the same. In this case we only focus on minimizing (18). From (18), we know $\max_{g\in G_2} Aloss$ is the function of $m(x)$. And because $\int_0^1 m(x)\,dx = 1$, let

$$\begin{aligned}\rho(m) &= \tau^2\left(\lambda_{\max}\left(\mathbf{K}(m)\mathbf{B}^{-2}(m)\right) + \nu\cdot tr\mathbf{B}^{-1}(m)\mathbf{D}(m)\mathbf{B}^{-1}(m)\right) + t\left(\int_0^1 m(x)\,dx - 1\right)\\ &= \rho_1(m) + \rho_2(m) + t\left(\int_0^1 m(x)\,dx - 1\right),\end{aligned}$$

where $\rho_1(m) = \tau^2\lambda_{\max}\left(\mathbf{K}(m)\mathbf{B}^{-2}(m)\right)$, $\rho_2(m) = \tau^2\nu\cdot tr\mathbf{B}^{-1}(m)\mathbf{D}(m)\mathbf{B}^{-1}(m)$. Using the Lagrange multipliers method, we get

$$\begin{cases}\partial_m\rho(m) = \partial_m\rho_1(m) + \partial_m\rho_2(m) + t\partial m = 0\\ \partial_t\rho(m) = \left(\int_0^1 m(x)\,dx - 1\right)\partial t = 0\end{cases}. \tag{A.8}$$

Because

$$\rho_2(m) = \tau^2\nu\cdot tr\mathbf{B}^{-1}(m)\mathbf{D}(m)\mathbf{B}^{-1}(m),$$

we have

$$\begin{aligned}\partial_m\rho_2(m) &= \tau^2\nu\cdot\partial_m tr\left[\mathbf{B}^{-1}(m)\mathbf{D}(m)\mathbf{B}^{-1}(m)\right]\\ &= \tau^2\nu\cdot tr\left\{-\mathbf{B}^{-1}(m)\mathbf{B}'(m)\mathbf{B}^{-1}(m)\mathbf{D}(m)\mathbf{B}^{-1}(m)\right.\\ &\quad\left.+\mathbf{B}^{-1}(m)\mathbf{D}'(m)\mathbf{B}^{-1}(m) - \mathbf{B}^{-1}(m)\mathbf{D}(m)\mathbf{B}^{-1}(m)\mathbf{B}'(m)\mathbf{B}^{-1}(m)\right\}\\ &= \tau^2\nu\left\{\pi^{-1}(x)\mathbf{f}^T(x)\mathbf{B}^{-2}(m)\mathbf{f}(x)\right.\\ &\quad\left.-\mathbf{f}^T(x)\left[\mathbf{B}^{-1}(m)\mathbf{D}(m)\mathbf{B}^{-2}(m) + \mathbf{B}^{-2}(m)\mathbf{D}(m)\mathbf{B}^{-1}(m)\right]\mathbf{f}(x)\right\}\\ &= \pi^{-1}(x)\mathbf{f}^T(x)\mathbf{N_w}\mathbf{f}(x) - \mathbf{f}^T(x)\mathbf{W_w}\mathbf{f}(x),\end{aligned}$$

where $\mathbf{N_w} = \tau^2\nu\mathbf{B}^{-2}(m)$ and $\mathbf{W_w} = \mathbf{B}^{-1}(m)\mathbf{D}(m)\mathbf{B}^{-2}(m) + \mathbf{B}^{-2}(m)\mathbf{D}(m)\mathbf{B}^{-1}(m)$. And

$$\begin{aligned}\rho_1(m) &= \tau^2\lambda_{\max}\left(\mathbf{K}(m)\mathbf{B}^{-2}(m)\right)\\ &= \tau^2\lambda_{\max}\left(\mathbf{B}^{-1}(m)\mathbf{K}(m)\mathbf{B}^{-1}(m)\right)\\ &= \tau^2\max_{\mathbf{d}\in\Re^p,\ \mathbf{d}\neq\mathbf{0}}\frac{\mathbf{d}\mathbf{B}^{-1}(m)\mathbf{K}(m)\mathbf{B}^{-1}(m)\mathbf{d}}{\mathbf{d}^T\mathbf{d}},\end{aligned}$$

where $\mathbf{d}$ is any vector in $\Re^p$, and we let vector $\mathbf{w} = \mathbf{B^{-1}}(\mathbf{m})\,\mathbf{d}$. Then $\mathbf{w} \in \Re^p$, $\mathbf{w} \neq \mathbf{0}$ if and only if $\mathbf{v} \in \Re^p$, $\mathbf{v} \neq \mathbf{0}$, since $\mathbf{K}$ is positive definite. Thus

$$\begin{aligned}\rho_1(m) &= \tau^2 \lambda_{\max}\left(\mathbf{B}^{-1}(m)\,\mathbf{K}(m)\,\mathbf{B}^{-1}(m)\right)\\ &= \tau^2 \max_{\mathbf{w}\in\Re^p,\ \mathbf{w}\neq\mathbf{0}} \frac{\mathbf{w}^T\mathbf{K}(m)\,\mathbf{w}}{\mathbf{w}^T\mathbf{B}^2(m)\,\mathbf{w}}\\ &= \tau^2 \max_{\|\mathbf{w}\|=1} \frac{\mathbf{w}^T\mathbf{K}(m)\,\mathbf{w}}{\mathbf{w}^T\mathbf{B}^2(m)\,\mathbf{w}}.\end{aligned}$$

Suppose that all the entries of matrices $\mathbf{B}(m)$ and $\mathbf{K}(m)$ are continuously differentiable functions of $m \in L_2(S)$. Then

$$\rho_1(m) = \tau^2 \max_{\|\mathbf{w}\|=1} \frac{\mathbf{w}^T\mathbf{K}(m)\,\mathbf{w}}{\mathbf{w}^T\mathbf{B}^2(m)\,\mathbf{w}},$$

is locally Lipschitz and its generalized gradient at m is

$$\partial\rho_1(m) = \tau^2 \cdot co\left\{\left[\frac{\mathbf{w}^T\mathbf{K}(m)\,\mathbf{w}}{\mathbf{w}^T\mathbf{B}^2(m)\,\mathbf{w}}\right]_m' : \mathbf{w}\in M(m)\right\},$$

where

$$M(m) = \left\{\mathbf{w} : \frac{\mathbf{w}^T\mathbf{K}(m)\,\mathbf{w}}{\mathbf{w}^T\mathbf{B}^2(m)\,\mathbf{w}} = \max_{\|\mathbf{w}\|=1}\frac{\mathbf{w}^T\mathbf{K}(m)\,\mathbf{w}}{\mathbf{w}^T\mathbf{B}^2(m)\,\mathbf{w}}\right\},$$

and

$$coA = \left\{\sum p_i a_i : p_i \geq 0, \sum p_i = 1, a_i \in A\right\},$$

is the convex hull of set A. For a vector $\mathbf{w} \in \Re^p$,

$$\begin{aligned}\left[\frac{\mathbf{w}^T\mathbf{K}(m)\,\mathbf{w}}{\mathbf{w}^T\mathbf{B}^2(m)\,\mathbf{w}}\right]_m' &= \frac{\mathbf{w}^T\mathbf{K}'(m)\,\mathbf{w}}{\mathbf{w}^T\mathbf{B}^2(m)\,\mathbf{w}}\\ &\quad - \frac{\left[\mathbf{w}^T\left[\mathbf{B}^2(m)\right]'\mathbf{w}\right]\left[\mathbf{w}^T\mathbf{K}(m)\,\mathbf{w}\right]}{\left[\mathbf{w}^T\mathbf{B}^2(m)\,\mathbf{w}\right]^2}\\ &= \frac{-2\mathbf{f}^T(x)\left[\mathbf{w}\mathbf{w}^T\mathbf{B}(m)\,\mathbf{w}^T\mathbf{K}(m)\,\mathbf{w} + \mathbf{w}\mathbf{w}^T\mathbf{B}(m)\,\mathbf{w}^T\mathbf{B}^2(m)\,\mathbf{w}\right]\mathbf{f}(x)}{\left[\mathbf{w}^T\mathbf{B}^2(m)\,\mathbf{w}\right]^2}\\ &\quad + \frac{2\mathbf{f}^T(x)\,\mathbf{w}\mathbf{w}^T\mathbf{f}(x)\,m(x)}{\left[\mathbf{w}^T\mathbf{B}^2(m)\,\mathbf{w}\right]^2}\\ &= -\mathbf{f}^T(x)\,\mathbf{M}_{\mathbf{w}1}\mathbf{f}(x) + \left[\mathbf{b}_{\mathbf{w}}^T\mathbf{f}(x)\right]^2 m(x),\end{aligned}$$

where $\mathbf{M}_{\mathbf{w}} = -\frac{2\left[\mathbf{w}\mathbf{w}^T\mathbf{B}(m)\mathbf{w}^T\mathbf{K}(m)\mathbf{w} + \mathbf{w}\mathbf{w}^T\mathbf{B}(m)\mathbf{w}^T\mathbf{B}^2(m)\mathbf{w}\right]}{\left[\mathbf{w}^T\mathbf{B}^2(m)\mathbf{w}\right]^2}$, $\mathbf{b}_{\mathbf{w}}^T = \sqrt{2}\frac{\mathbf{w}^T}{\mathbf{w}^T\mathbf{B}^2(m)\mathbf{w}}$. So (A.8) equals to

$$\begin{cases}\pi^{-1}(x)\,\mathbf{f}^T(x)\,\mathbf{N}_{\mathbf{w}}\mathbf{f}(x) - \mathbf{f}^T(x)\,\mathbf{W}_{\mathbf{w}}\mathbf{f}(x) - \mathbf{f}^T(x)\sum_{i=1}^N p_i\mathbf{M}_{\mathbf{w}_i}\mathbf{f}(x)\\ \qquad + \sum_{i=1}^N p_i\left[\mathbf{b}_{\mathbf{w}_i}^T\mathbf{f}(x)\right]^2 m(x) + t = 0\\ \int_0^1 m(x)\,dx - 1 = 0\end{cases}.$$

Consequently, there exist a constant symmetric matrix $\mathbf{M}$, a constant positive semi-definite matrix $\mathbf{U}$, a constant symmetric positive definite matrix $\mathbf{N}$ and a constant t such that

$$\mathbf{f}^T(x)\,\mathbf{U}\mathbf{f}(x)\,m(x) + \pi^{-1}(x)\,\mathbf{f}^T(x)\,\mathbf{N}\mathbf{f}(x) - \mathbf{f}^T(x)\,\mathbf{M}\mathbf{f}(x) + t = 0, \tag{A.9}$$

for all x such that $m(x) > 0$. Since $\mathbf{f}^T(x)\,\mathbf{U}\mathbf{f}(x) > 0$, by solving (A.9), we have (20).□

References

Alpert, B.K. (1992).Wavelets and Other Bases for Fast Numerical Linear Algebra, in: Wavelets. *A Tutorial in Theory and Applications*, Ed: Chui, C.K., Boston: Academic Press, 181-216.

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.

Brown, P.J., Fearn, T., and Vannucci, M. (2001). Bayesian Wavelet Regression on Curves With Application to a Spectroscopic Calibration Problem. *Journal of the American Statistical Association*, 96, No. 454, 398-408.

Capilla, C. (2005). Application of the Haar Wavelet Transform to Detect Microseismic Signal Arrivals. *Journal of Applied Geophysics*, 59, No. 1, 36-46.

Daubechies, I. (1992). Ten Lectures on Wavelets. *CBMS-NSF Regional Conference Series in Applied Mathematics.* Lectures delivered at the CBMS conference organized in June 1990 by the Mathematics Department at the University of Lowell, Massachusetts, US. Society for Industrial and Applied Mathematics.

Haar, A. (1910). Zur theorie der orthogonalen funktionensysteme. *Mathematische Annalen*, 69, 331-371.

Herzberg, A.M., and Traves, W.N. 1994. An Optimal Experimental Design for the Haar Regression Model. *Canadian Journal of Statistics*, 22, 357-364.

Murat, K., and Necati, A. (2006). Wavelet Regression Technique for Streamflow Prediction. *Journal of Applied Statistics*, 33, 943-960.

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

Robert, D.N., and Richard, G.B. (1999). Wavelet-Based Transformations for Nonlinear Signal Processing. *IEEE Transactions on Signal Processing*, 47, No. 7, 1852-1865.

Strang, G. (1989). Wavelets and Dilation Equations: A Brief Introduction. *SIAM Review*, 31, 614-627.

Tian, Y.G., Herzberg, A.M. (2006A). D-optimal Designs for a Combined Simple Linear and Haar-Wavelet Regression Model. *Communications in Statistics - Theory and Methods*. 27, 1085-1091.

Tian, Y.G., Herzberg, A.M. (2006B). A-minimax and D-minimax Robust Optimal Designs for Approximately Linear Haar-wavelet Models. *Computational Statistics and Data Analysis*, 50, No. 10, 2942-2951.

Xie, M.-Y. (2002). D-optimal designs based on elementary intervals for b-adic Haar wavelet regression models. *Monte Carlo and quasi-Monte Carlo methods*. Eds: Niederreiter, H., Spanier, J. Berlin: Springer-Verlag, 523-535.

Table 1: Convergence Time (in millisecond), $\delta = \frac{1}{2^{m+1}*1000}$

m	$\frac{\sigma^2}{n\tau^2}$	$\pi(x) = x^2$	$\pi(x) = \exp(-x)$
3	neglectable	16	16
3	1	15	15
4	neglectable	32	78
4	1	31	63
5	neglectable	141	266
5	1	62	63
6	neglectable	609	344
6	1	312	188

Table 2: Coefficients in (21)

	$\pi(x) = x$				$\pi(x) = \frac{1}{x}$			
ν	a_1	a_2	b_1	b_2	a_1	a_2	b_1	b_2
0.1	1.5085	0.0027	1.0560	0.3813	0.7471	0.0602	1.4887	0.2961
1	1.5111	0.0080	0.9966	0.3027	0.9111	0.7746	2.3387	1.4105
5	1.5485	0.0166	1.0532	0.3181	1.4802	3.4060	6.7093	7.1824

Table 3: Coefficients in (21), continued

	$\pi(x)=e^x$				$\pi(x)=e^{-x}$			
ν	a_1	a_2	b_1	b_2	a_1	a_2	b_1	b_2
0.1	1.1764	0.0638	0.9323	0.1245	0.9812	0.0808	1.4029	0.1311
1	1.0719	−0.0657	0.9033	0.0583	2.3518	1.1377	3.0309	0.8921
5	4.6228	4.4457	3.6020	5.7124	6.2540	4.1961	8.0297	3.2678

Table 4: Losses and Relative Efficiencies

	$\pi(x)=x$			$\pi(x)=\frac{1}{x}$		
ν	$MinAloss$	$Uloss$	eff	$MinAloss$	$Uloss$	eff
0.1	0.9456	1.9138	2.024	0.0928	0.1000	1.077
1	8.1861	19.1376	2.338	0.9080	0.9995	1.101
5	36.6151	95.6881	2.613	4.0400	4.9975	1.237

Table 5: Losses and Relative Efficiencies, continued

	$\pi(x)=e^x$			$\pi(x)=e^{-x}$		
ν	$MinAloss$	$Uloss$	eff	$MinAloss$	$Uloss$	eff
0.1	0.1244	0.1265	1.0164	0.3373	0.3436	1.0185
1	1.2455	1.2645	1.0153	3.3237	3.4357	1.0337
5	6.0239	6.3226	1.0496	15.6874	17.1824	1.0953

Table 6: Discretized Design Points

$\pi(x)=x$			$\pi(x)=\frac{1}{x}$			$\pi(x)=e^x$			$\pi(x)=e^{-x}$		
$\nu=0.1$	$\nu=1$	$\nu=5$	$\nu=0.1$	$\nu=1$	$\nu=5$	$\nu=0.1$	$\nu=1$	$\nu=5$	$\nu=0.1$	$\nu=1$	$\nu=5$
0.041	0.050	0.062	0.067	0.056	0.118	0.045	0.044	0.117	0.057	0.042	0.025
0.109	0.121	0.135	0.203	0.178	0.229	0.135	0.133	0.231	0.168	0.133	0.080
0.171	0.190	0.204	0.340	0.317	0.355	0.223	0.222	0.312	0.282	0.234	0.144
0.243	0.258	0.272	0.478	0.484	0.509	0.312	0.310	0.380	0.397	0.355	0.227
0.310	0.325	0.339	0.563	0.557	0.543	0.401	0.400	0.439	0.509	0.508	0.505
0.376	0.392	0.405	0.639	0.619	0.580	0.488	0.490	0.494	0.595	0.574	0.544
0.443	0.459	0.471	0.717	0.692	0.622	0.600	0.601	0.694	0.681	0.646	0.589
0.543	0.588	0.617	0.796	0.769	0.671	0.716	0.716	0.805	0.770	0.726	0.639
0.758	0.770	0.784	0.876	0.855	0.730	0.830	0.820	0.892	0.860	0.817	0.699
0.924	0.927	0.931	0.959	0.945	0.816	0.944	0.943	0.966	0.953	0.928	0.786

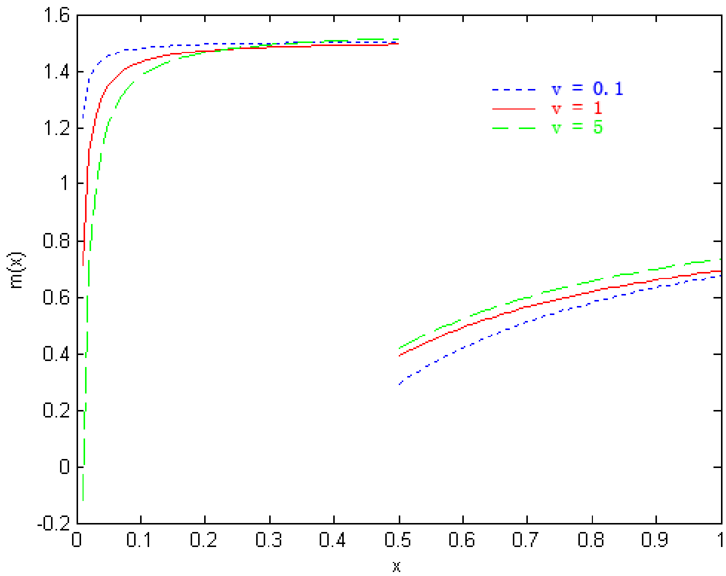

Figure 1: Design density $m(x)$ when $\pi(x) = x$.

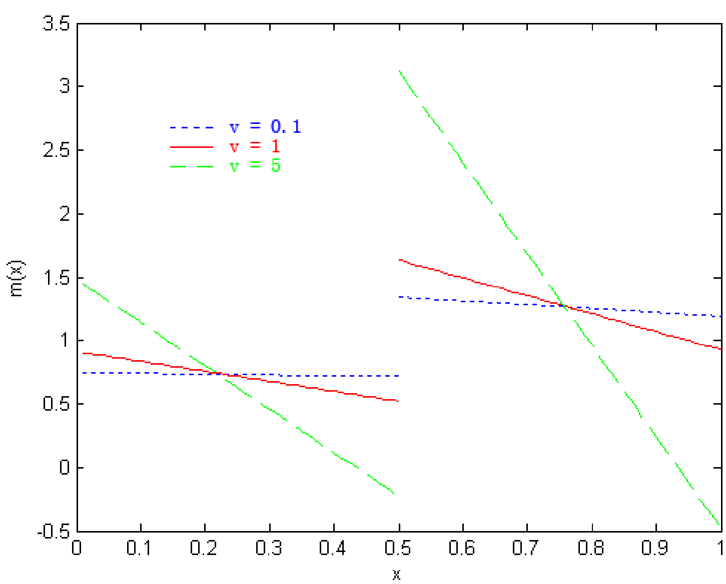

Figure 2: Design density $m(x)$ when $\pi(x) = \frac{1}{x}$.

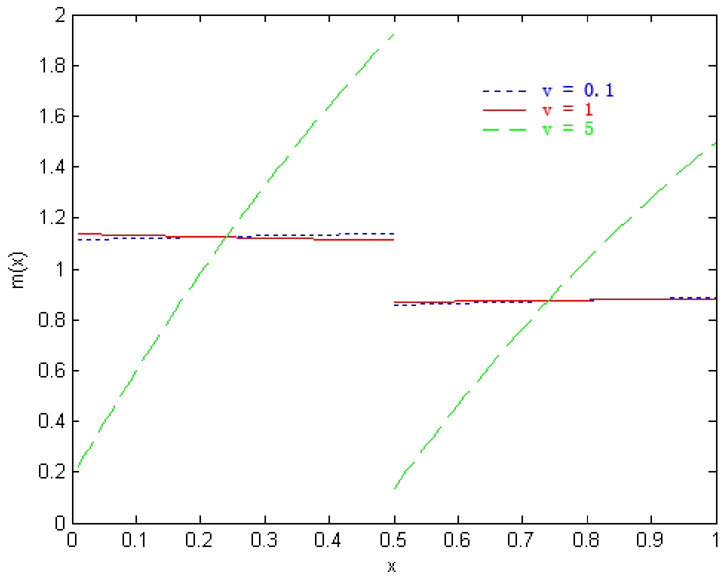

Figure 3: Design density $m(x)$ when $\pi(x) = e^{x}$.

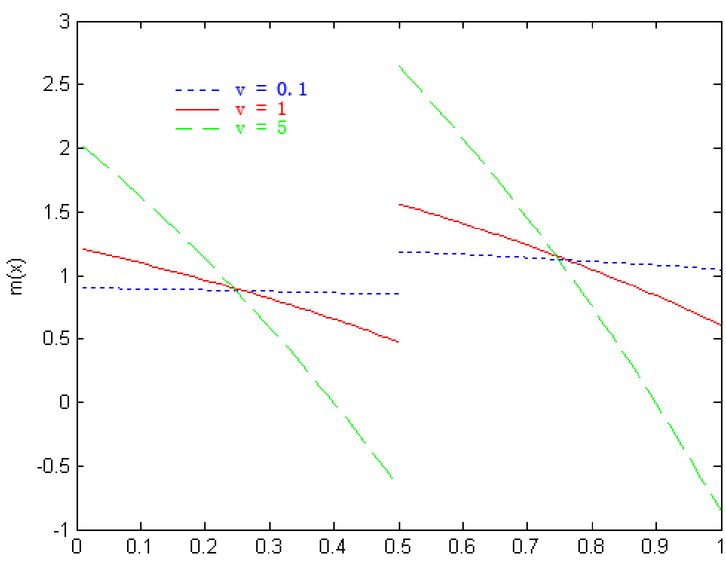

Figure 4: Design density $m(x)$ when $\pi(x) = e^{-x}$.

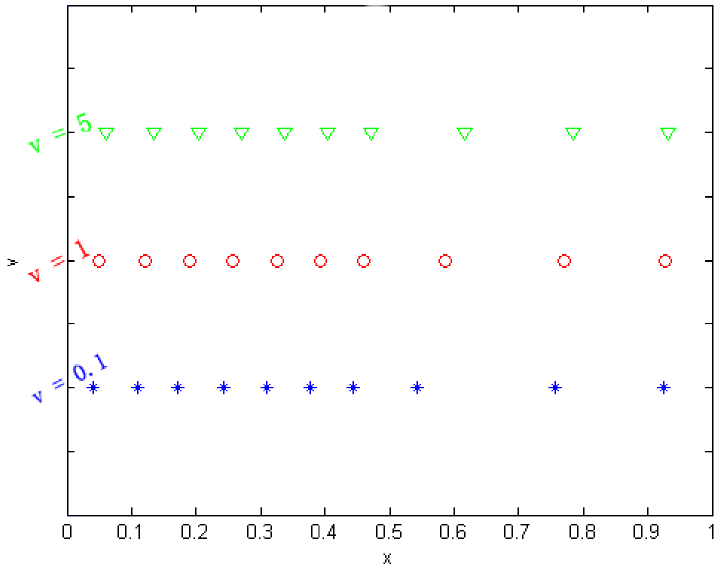

Figure 5: Discrete design points x_i for various ν when $\pi(x) = x$.

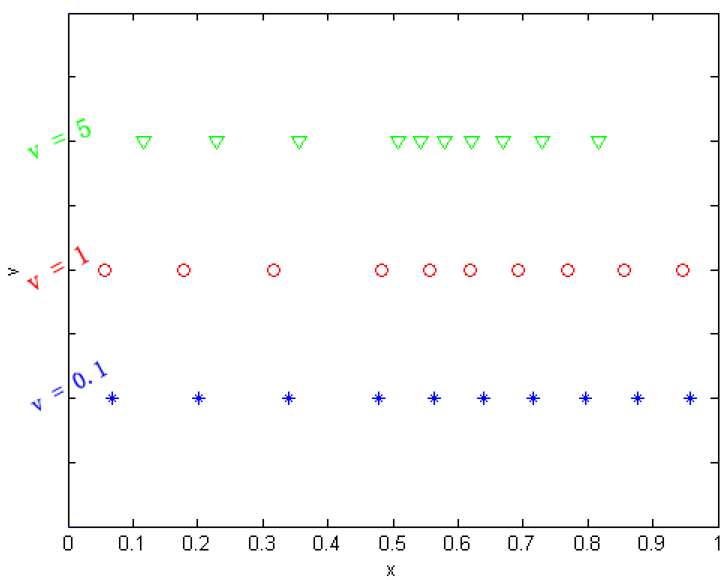

Figure 6: Discrete design points x_i for various ν when $\pi(x) = \frac{1}{x}$.

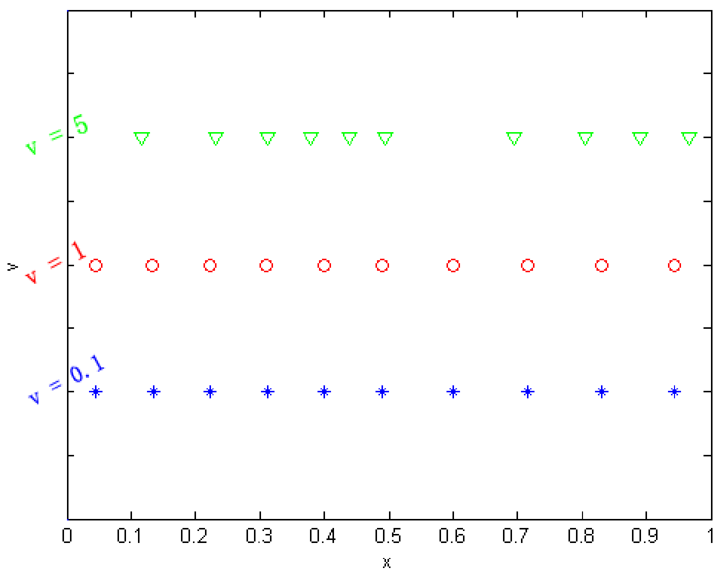

Figure 7: Discrete design points x_i for various ν when $\pi(x) = e^x$.

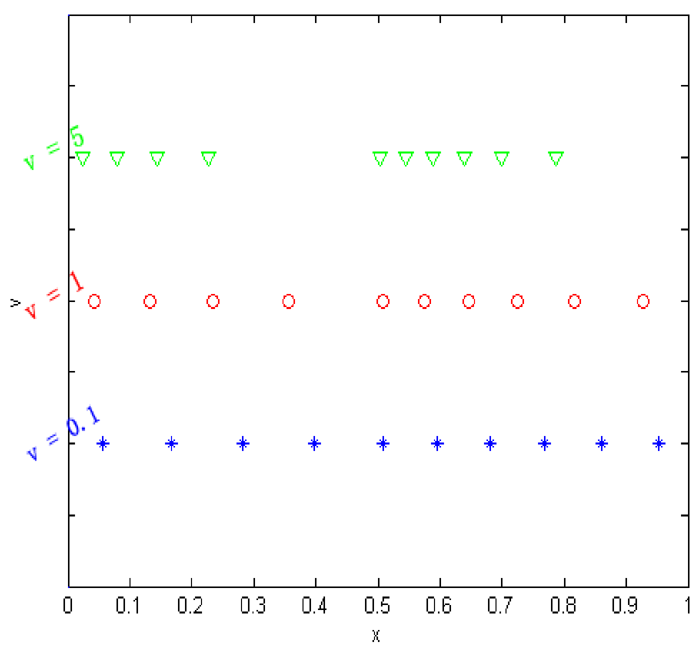

Figure 8: Discrete design points x_i for various ν when $\pi(x) = e^{-x}$.