Robust Designs for Approximate Regression Models With Correlated Errors

Douglas P. Wiens
Department of Mathematical and Statistical Sciences
University of Alberta, Edmonton, Alberta, Canada T6G 2G1
doug.wiens@ualberta.ca

KEY WORDS Approximately linear regression response; Autoregressive errors; Infinitesimal approach; Integrated Mean Squared Error; Minimax; Nearest neighbour; Regression design

Abstract

We summarize recent findings on the construction of designs, to be used in experiments for which regression is the anticipated method of analysis. The designs are to be robust against departures from the assumed linear response and/or departures from the assumption of uncorrelated errors. Included is a discussion of an "infinitesimal" approach to design. In this approach one aims to minimize the determinant of the mean squared error matrix of the regression estimates, subject to the satisfaction of a robustness constraint. The constraint is quantified in terms of boundedness of the Gateaux derivative of this determinant, in the direction of a contaminating response function or autocorrelation structure.

1. INTRODUCTION

We summarize recent findings on the construction of designs, to be used in experiments for which regression is the anticipated method of analysis. The designs are to be robust against departures from the assumed linear response and/or departures from the assumption of uncorrelated errors. Departures of this first type are modelled in Section 2. In Section 3 we consider regression models with correlated errors. It turns out that, generally speaking, robustness against both types of departures is attained by applying, in an appropriate order, the points from a design which is robust against departures of the first type. We illustrate this in three cases. The first allows for a very broad class of autocorrelation structures; the second for AR(1) errors. In the third case we discuss an "infinitesimal" approach to design. In this approach one aims to minimize the determinant of the mean squared error matrix of the regression estimates, subject to the satisfaction of a robustness constraint. The constraint is quantified in terms of boundedness of the Gateaux derivative of this determinant, in the direction of a contaminating response function or autocorrelation structure. Each approach is illustrated in an example.

2. MODELLING DEPARTURES FROM LINEARITY

The regression model which we envisage is one for which the usual parameter estimates are biased.

Example 1: An experiment of interest at an environmental research station in Alberta involved the treatment of wastewater with preset amounts of x_1 = ozone and x_2 = chlorine; the response variable was a measure Y of the remaining pollutants. For various reasons, including a carryover effect, the reponses were correlated over time, but in a manner difficult to model. As well, the experimenters assumed a linear (in the parameters) model relating the response to the covariates, while being aware that this was only a working approximation. Thus, they desired a design robust against these sources of model misspecification.

More generally, suppose that the experimenter is to take observations on a random variable Y obeying the 'approximately linear' model

$$Y(\mathbf{x}) = \mathbf{z}^T(\mathbf{x})\boldsymbol{\theta} + f(\mathbf{x}) + \epsilon, \qquad (1)$$

for some p-dimensional parameter vector $\boldsymbol{\theta}$ and regressors $\mathbf{z} = \mathbf{z}(\mathbf{x})$, errors ϵ with common variance σ^2, and an unknown contaminant $f(\mathbf{x})$ representing uncertainty about the exact nature of the regression response. For example, the elements of $\mathbf{z}$ could be low-degree monomials in the elements of $\mathbf{x}$, with $f(\mathbf{x})$ being a multinomial of higher degree. Only $\boldsymbol{\theta}$ is estimated, and then $E[Y|\mathbf{x}]$ is estimated by $\hat{Y}(\mathbf{x}) = \mathbf{z}^T(\mathbf{x})\hat{\boldsymbol{\theta}}$, so that $\hat{Y}(\mathbf{x})$ can be highly biased both for $\mathbf{z}^T(\mathbf{x})\boldsymbol{\theta}$ and for $E[Y|\mathbf{x}]$. Protection is sought against this bias, and against errors due to random variation.

To obtain (1) we first suppose that the experimenter is to take n observations on a random variable Y whose mean is thought to vary, in an approximately linear manner, with regressors $\mathbf{z}(\mathbf{x})$: $E[Y|\mathbf{x}] \approx \mathbf{z}^T(\mathbf{x})\boldsymbol{\theta}$. The sites $\mathbf{x}_i$ are chosen from $\mathcal{S}$, a q-dimensional design space with volume. We define the "true" value of $\boldsymbol{\theta}$ by requiring the linear approximation to be most accurate in the L^2-sense:

$$\boldsymbol{\theta} := \arg\min_{\mathbf{t}} \int_{\mathcal{S}} (E[Y|\mathbf{x}] - \mathbf{z}^T(\mathbf{x})\mathbf{t})^2 d\mathbf{x}.$$

We then define $f(\mathbf{x}) = E[Y|\mathbf{x}] - \mathbf{z}^T(\mathbf{x})\boldsymbol{\theta}$ and $\epsilon(\mathbf{x}) = Y(\mathbf{x}) - E[Y|\mathbf{x}]$, so that (1) holds. These definitions of $\boldsymbol{\theta}$ and of f together imply that

$$\int_{\mathcal{S}} \mathbf{z}(\mathbf{x}) f(\mathbf{x}) d\mathbf{x} = \mathbf{0}. \tag{2}$$

In order that errors due to bias will not swamp those due to variance, we shall also assume a bound

$$\int_{\mathcal{S}} f^2(\mathbf{x}) d\mathbf{x} \leq \eta^2 < \infty. \tag{3}$$

Denote by $\mathcal{F}$ the class of functions f satisfying (2) and (3).

2.1 Loss functions

An exactly implementable design will correspond to a *design measure* ξ placing mass n^{-1} at each of $\mathbf{x}_1, ..., \mathbf{x}_n$. Below, we exhibit the moments of the least squares estimator $\hat{\boldsymbol{\theta}}$ under such a design. As is common in design theory, we then broaden the class of allowable measures to the class Ξ of all probability measures on $\mathcal{S}$. We find optimal designs in this class and approximate them, as necessary, prior to implementation.

When the model $E[Y|\mathbf{x}] = \mathbf{z}^T(\mathbf{x})$ is fitted although the true model is (1), $\hat{\boldsymbol{\theta}}$ is biased. With $\mathbf{b} := \int_{\mathcal{S}} \mathbf{z}(\mathbf{x}) f(\mathbf{x})\, \xi(d\mathbf{x})$ and $\mathbf{B} := \int_{\mathcal{S}} \mathbf{z}(\mathbf{x})\mathbf{z}^T(\mathbf{x})\, \xi(d\mathbf{x})$ assumed non-singular, the bias is $E[\hat{\boldsymbol{\theta}}] - \boldsymbol{\theta} = \mathbf{B}^{-1}\mathbf{b}$ and the mean squared error matrix is

$$\begin{aligned} \mathrm{MSE}(f,\xi) &= E\left\{(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})^T\right\} \\ &= (\frac{\sigma^2}{n})\mathbf{B}^{-1} + \mathbf{B}^{-1}\mathbf{b}\mathbf{b}^T\mathbf{B}^{-1}. \end{aligned}$$

We consider the loss functions $\mathcal{L}_Q$ = integrated MSE of the fitted responses $\hat{Y}(\mathbf{x})$, $\mathcal{L}_D$ = determinant of the MSE matrix and $\mathcal{L}_A$ = trace of the MSE matrix. These correspond to the classical notions of Q-, D- and A-optimality, and so we adopt the same nomenclature. (The term Q-optimality seems to be due to Fedorov (1972); Studden (1977) and others have used instead the term I-optimality.) Explicit descriptions of these loss functions are, with $\mathbf{A} := \int_{\mathcal{S}} \mathbf{z}(\mathbf{x})\mathbf{z}^T(\mathbf{x}) d\mathbf{x}$, given by

$$\begin{aligned} \mathcal{L}_Q(f,\xi) &= \int_S E\left[\left\{\hat{Y}(\mathbf{x}) - E(Y|\mathbf{x})\right\}^2\right] d\mathbf{x} \\ &= \mathrm{tr}\left(\mathrm{MSE}(f,\xi)\mathbf{A}\right) + \int_S f^2(\mathbf{x}) d\mathbf{x} \\ &= \left(\frac{\sigma^2}{n}\right) \mathrm{tr}\left(\mathbf{B}^{-1}\mathbf{A}\right) \\ &\quad + \mathbf{b}^T\mathbf{B}^{-1}\mathbf{A}\mathbf{B}^{-1}\mathbf{b} + \int_S f^2(\mathbf{x}) d\mathbf{x}, \\ \mathcal{L}_D(f,\xi) &= \det\left\{\mathrm{MSE}(f,\xi)\right\} \\ &= \left(\frac{\sigma^2}{n}\right)^p \frac{1}{|\mathbf{B}|}\left\{1 + \frac{n}{\sigma^2}\mathbf{b}^T\mathbf{B}^{-1}\mathbf{b}\right\}, \\ \mathcal{L}_A(f,\xi) &= \mathrm{tr}\left\{\mathrm{MSE}(f,\xi)\right\} \\ &= \left(\frac{\sigma^2}{n}\right) \mathrm{tr}\mathbf{B}^{-1} + \mathbf{b}^T\mathbf{B}^{-2}\mathbf{b}. \end{aligned}$$

We aim to construct designs to minimize the maximum (over $\mathcal{F}$) value of the loss.

2.2 Example

For the multiple linear regression model (with constant):

$$\mathbf{z}(\mathbf{x}) = (1, x_1, ..., x_q)^T, \tag{4}$$

with $\mathbf{x}$ ranging over a spherical design space

$$||\mathbf{x}|| \leq r, \tag{5}$$

the Q-optimal *minimax* density was obtained by Huber (1975) for $q = 1$ and by Wiens (1990) for $q > 1$. The minimax procedure involves first maximizing the loss over f, subject to (2) and (3), and then minimizing over all probability measures ξ. For second and higher order models the "min" half of the minimax problem is quite intractable, and so the problem remains open. Heo, Schmuland and Wiens (2001) derived approximately minimax procedures which yield design densities for general response functions and arbitrary design spaces.

The designs can then be implemented by placing design points at the quantiles of the continuous design measure ξ_* in the univariate case, or at multivariate analogues of the quantiles. Another option is to match up $O(n)$ of the moments. The latter approach was taken by Heo, Schmuland and Wiens (2001) - see Figure

1 for an implementation in the context of Example 1. A further possibility is to take a finite (but dense) design space and to then optimally choose the design points directly, through simulated annealing - see Fang and Wiens (2000), Oyet and Wiens (2000) and Zhou (2001b,c).

Using any of these approaches one obtains an implementable design ξ_n converging weakly to ξ_*, so that the optimality is preserved, asymptotically. A rough description of these implementations is that they replace the point masses of the "classically" optimal design (which assumes the fitted model to be exactly correct) with clusters of design points at nearby but distinct locations.

3. DESIGNING FOR DEPARTURES FROM INDEPENDENCE

The above approach can be broadened to allow for correlated errors. We discuss three approaches.

3.1 Omnibus correlations

Take

$$\mathrm{CORR}(\boldsymbol{\varepsilon}) = (1-\alpha)\,\mathbf{I} + \alpha\mathbf{Q}$$

for $\alpha \in [0,1)$ and $\mathbf{Q}$ an arbitrary autocorrelation matrix. Equivalently (assuming the autocorrelations to be absolutely summable) the spectral density of the error process is

$$g(\omega) = (1-\alpha)\,g_0(\omega) + \alpha g_1(\omega)$$

where $g_0(\omega)$ is uniform (corresponding to uncorrelated errors) and $g_1(\omega)$ is arbitrary.

In the minimax problem we now maximize over both departures from linearity and over $g_1(\omega)$, and then minimize over the design measure and the order in which the design points are implemented. Then (under several technical conditions):

Theorem 1 *(Wiens and Zhou 1996) An asymptotically minimax design mechanism is attained by either*
(i) randomly sampling design points from the measure ξ_* which is minimax for the approximate regression model with independent errors; or
(ii) randomly permuting the points from a design ξ_n, where ξ_n converges weakly to ξ_*.

3.2 Autoregressive errors

If more structure is imposed on the autocorrelations, then more precise results are possible. Suppose that the errors follow an AR(1) model, so that

$$\mathrm{COV}(\boldsymbol{\varepsilon}) = \frac{\sigma^2}{1-\rho^2}\mathbf{P},\ \ P_{i,j} = \rho^{|i-j|},\ \ 0 \le |\rho| < 1$$

and that the estimate is the BLUE: $\hat{\boldsymbol{\theta}} = \left(\mathbf{X}^T\hat{\mathbf{P}}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}^T\hat{\mathbf{P}}^{-1}\mathbf{y}$. Take the model defined by (4) and (5). Then:

Theorem 2 *(Wiens and Zhou 1999) Fix the sign of ρ. Suppose that the empirical distributions $\xi_n(\mathbf{u},\mathbf{v})$ of the consecutive design points $(\mathbf{x}_i, \mathbf{x}_{i+1})$, $i = 1,\ldots,n-1$ satisfy:*
1) The marginal distributions of ξ_n converge weakly to ξ_* (the optimal design for uncorrelated errors); and
2) $E_{\xi_n}\{||\mathbf{U} + sign(\rho)\mathbf{V}||^2\} \to 0$ as $n \to \infty$.
Then the sequence of designs $\{\xi_n\}_{n=1}^{\infty}$ is asymptotically minimax.

When $\rho < 0$, requirement 2) of Theorem 2 states that the average squared distance between consecutive design points is to tend to 0. When $\rho > 0$, the average of the squared norms of the pairwise averages of consecutive design points is to tend to 0. Thus, if an optimal design for negative ρ has support points $\{\mathbf{x}_i\}_{i=1}^n$, the design with support points $\{(-1)^i\mathbf{x}_i\}_{i=1}^n$ is optimal for positive ρ.

For negative ρ, one can implement the ordering by defining $\mathbf{x}_0 = \mathbf{0}$, and then taking $\mathbf{x}_{i+1}$ as the nearest (remaining) neighbour of $\mathbf{x}_i$, $i = 1,\ldots,n-1$. Prior to the ordering, the $\mathbf{x}$'s can be a random sample from ξ_*. See Figure 2 for examples of implementations based on ξ_* of Figure 1.

3.3 Infinitesimal approach

Suppose that $\mathrm{CORR}(\boldsymbol{\varepsilon}) = (1-\alpha)\,\mathbf{I} + \alpha\mathbf{Q}$ and let $\mathcal{D}((1-\alpha)\,\mathbf{I} + \alpha\mathbf{Q})$ be the corresponding determinant of the covariance matrix of the LSE $\hat{\boldsymbol{\theta}}$. Define the *change-of-variance function*

$$\mathrm{CVF}(\xi,\mathbf{Q}) = \frac{\frac{d}{d\alpha}\mathcal{D}((1-\alpha)\,\mathbf{I} + \alpha\mathbf{Q})_{|\alpha=0}}{\mathcal{D}(\mathbf{I})},$$

measuring the instantaneous change in the variance as one moves from independence, in the direction of $\mathbf{Q}$. The *change-of-variance sensitivity* in a class $\mathcal{Q}$ of autocorrelation matrices is

$$\mathrm{CVS}(\xi,\mathcal{Q}) = \sup_{\mathbf{Q}\in\mathcal{Q}} \mathrm{CVF}(\xi,\mathbf{Q}).$$

A design is *most V-robust* for $\mathcal{Q}$ if it minimizes $\mathcal{D}(\mathbf{I})$ in the class of all designs with

$\mathrm{CVS}(\xi, \mathcal{Q}) \leq \alpha$, with α being the minimum such bound. These designs were considered by Wiens and Zhou (1997) . Examples are given in Figure 3 for multiple linear regression ($q = 2$) over the unit square and classes $\mathcal{Q}_1$ and $\mathcal{Q}_2$ corresponding to MA(1) processes with, respectively, positive and negative lag-1 correlations bounded away from 0. See Zhou (2001a) for extensions to approximately linear models (controlling an analogously defined *change-of-bias function* as well) with IMSE loss.

REFERENCES

Fang, Z., and Wiens, D.P. (2000), "Integer-valued, Minimax Robust Designs for Estimation and Extrapolation in Heteroscedastic, Approximately Linear Models," *Journal of the American Statistical Association*, 95, 807-818.

Fedorov, V.V. (1972), *Theory of Optimal Experiments*, New York: Academic Press.

Heo, G., Schmuland, B., and Wiens, D.P. (2001); "Restricted Minimax Robust Designs for Misspecified Regression Models," *The Canadian Journal of Statistics*, 29, 117-128.

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

Oyet, A.J. and Wiens, Wiens, D.P. (2000); "On Exact Minimax Wavelet Designs Obtained by Simulated Annealing,", submitted.

Studden, W.J. (1977), "Optimal Designs for Integrated Variance in Polynomial Regression," *Statistical Decision Theory and Related Topics II*, ed. Gupta, S.S. and Moore, D.S. New York: Academic Press, pp. 411-420.

Wiens, D.P. (1990), "Robust, Minimax Designs for Multiple Linear Regression," *Linear Algebra and Its Applications, Second Special Issue on Linear Algebra and Statistics*; 127, 327 - 340.

Wiens, D.P., and Zhou, J. (1996), "Minimax Regression Designs for Approximately Linear Models With Autocorrelated Errors," *Journal of Statistical Planning and Inference*, 55, 95-106.

Wiens, D.P., and Zhou, J. (1997), "Robust Designs Based on the Infinitesimal Approach," *Journal of the American Statistical Association*, 92, 1503-1511.

Wiens, D.P., and Zhou, J. (1999), "Minimax Designs for Approximately Linear Models with AR(1) Errors," *The Canadian Journal of Statistics*, 27, 781-794.

Zhou, J. (2001), "Two Robust Design Approaches for Linear Models with Correlated Errors", *Statistica Sinica*, 11, 261-272.

Zhou, J. (2001), "Integer-valued, Minimax Robust Designs for Approximately Linear Models With Correlated Errors," *Communications in Statistics*, in press.

Zhou, J. (2001), "A Robust Criterion for Experimental Designs for Serially Correlated Observations", *Technometrics*, in press.

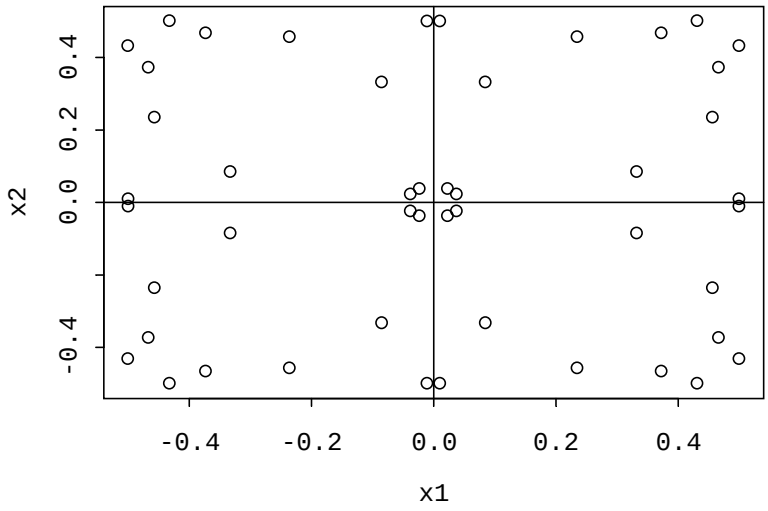

Figure 1: Robust design for $E[Y|\mathbf{x}] \approx$ second order in x_1 = ozone, x_2 = chlorine; $n = 48$. Covariates transformed to unit square. Note clusters of (distinct) design points near the classically optimal, replicated locations (corners and centre): bias reduction vs. variance reduction.

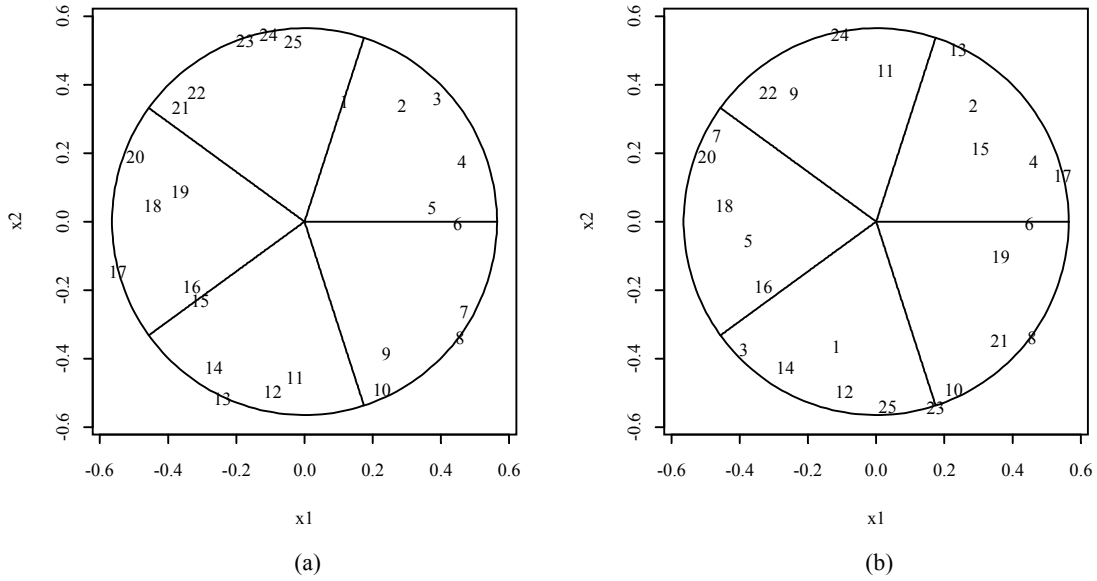

Figure 2: Designs for first order linear regression and AR(1) errors with (a) $\rho < 0$ (b) $\rho > 0$; $n = 25$. The design points are randomly chosen from ξ_* and then ordered using the nearest neighbour algorithm.

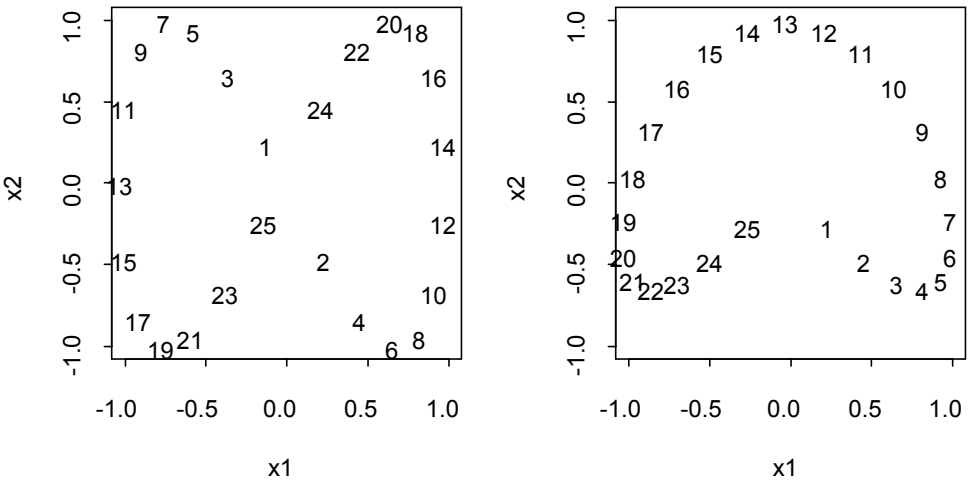

Figure 3: Most V-robust designs for first order linear regression and MA(1) errors with positive (left plot) and negative (right plot) lag-1 autocorrelations bounded away from 0; $n = 25$, $\mathcal{S} = [-1, 1] \times [-1, 1]$. Indices of ordered design points are plotted.