Minimum Variance Unbiased Weights and Designs for Biased Regression Models

Douglas P. Wiens University of Alberta

This paper has appeared in the Proceedings of the 3rd St. Petersburg Workshop on Simulation; June 28-July 3 1998.

We exhibit regression designs and weights, for estimation and for extrapolation, which are robust against incorrectly specified regression responses and possible error heteroscedasticity. Examples are given for polynomial and wavelet regression models. The results apply to Generalized M-estimation as well as to least squares estimation.

---

1 GENERAL THEORY

In this article we report some recent findings concerning the interplay between the choice of design points for regression based data analyses, and the weights used in Weighted Least Squares (WLS) or Generalized M-estimation. The regression model which we envisage is one for which the Ordinary Least Squares (OLS) estimates are biased. Furthermore, we allow for the possibility of error heteroscedasticity. Specifically, suppose that the experimenter is to take observations on a random variable Y obeying the **approximately linear** model

$$Y(x) = z^T(x)\mu + f(x) + ²; \tag{1}$$

for some p-dimensional parameter vector μ and regressors $z(x)$, and an unknown contaminant $f(x)$ representing uncertainty about the exact nature of the regression response. For example, the elements of z could be low-degree monomials in the elements of x, with $f(x)$ being a polynomial of higher degree. Only the linear coefficients are estimated - $\hat{Y}(x) = z^T(x)\hat{\mu}$ - so that $\hat{Y}(x)$ can be a highly biased estimate of $E[Y j x]$. Protection is sought against both this bias, and errors due to random variation and possible heteroscedasticity.

Our approach is to first determine designs which, for fixed weights, will minimize the maximum bias of $\hat{Y}(x)$ for $z^T(x)\mu$ as $E[Y j x]$ ranges over an L_2 neighbourhood of linear functions. The weights are then chosen to minimize (a function of) the covariance matrix of $\hat{\mu}$.

We suppose that the experimenter is to take n uncorrelated observations Y_i following the relationship (1). The sites x_i are chosen from a design space $S \subset R^q$ with volume $\int_S dx = \Omega^{-1}$. The response error $f(x)$ and the random errors $\varepsilon(x)$ satisfy

$$\int_S z(x) f(x) dx = 0, \qquad \int_S f^2(x) dx \le \eta^2, \tag{2}$$

$$E[\varepsilon(x)] = 0, \quad VAR[\varepsilon(x)] = \sigma^2 g(x), \quad \int_S g^2(x) dx \le \Omega^{-1}. \tag{3}$$

The first conditions of each of (2) and (3) correspond to defining θ, $f(x)$ and $\varepsilon(x)$ by $\theta = \arg\min_t \int_S (E[Y|x] - z^T(x)t)^2 dx$, $f(x) = E[Y|x] - z^T(x)\theta$, and $\varepsilon(x) = Y(x) - E[Y|x]$. Some condition like the second in (2) is necessary in order that errors due to bias not swamp those due to variance. The second and third conditions of (3) amount to defining $\sigma^2 = \sup_g \left(\int_S VAR^2[\varepsilon(x)]\Omega dx\right)^{1/2}$. Neither σ^2 nor η^2 need be known to the experimenter in order for our results to be applied.

We propose to estimate θ by least squares, possibly weighted with non-negative weights $w(x)$. Let ξ be the design measure, i.e. the empirical distribution function of $\{x_1, \ldots, x_n\}$. Define matrices A, B, D and a vector b by $A = \int_S z(x)z^T(x)dx$, $B = \int_S w(x)z(x)z^T(x)d\xi(x)$, $D = \int_S w^2(x)g(x)z(x)z^T(x)d\xi(x)$ and $b = \int_S w(x)z(x)f(x)d\xi(x)$. Assume that A and B are nonsingular and define $H = B^{-1}AB^{-1}$. In this notation, the mean vector and covariance matrix of the WLS estimate are $E[\hat\theta] - \theta = B^{-1}b$ and $COV[\hat\theta] = (\sigma^2/n)B^{-1}DB^{-1}$. We estimate $E[Y|x]$ by $\hat Y(x) = z^T(x)\hat\theta$, and consider the resulting Integrated Mean Squared Error $(IMSE) = \int_S E[(\hat Y(x) - E[Y|x])^2]dx$. This splits into terms due solely to estimation bias, estimation variance, and model misspecification:

$$IMSE(f, g, w, \xi) = ISB(f, w, \xi) + IV(g, w, \xi) + \int_S f^2(x),$$

where the Integrated Squared Bias (ISB) and Integrated Variance (IV) are

$$ISB(f, w, \xi) = \int_S \left(E[\hat Y(x) - z^T(x)\theta]\right)^2 dx = b^T H b,$$
$$IV(g, w, \xi) = \int_S VAR[\hat Y(x)]dx = \frac{\sigma^2}{n} \cdot trace(HD).$$

Although our discussion is carried out for least squares estimation, the results apply as well to Mallows-type Generalized M-Estimation. The reason is that the asymptotic bias and covariance structure of the GM-estimator are as given above, up to (for the covariance matrix) a constant multiple independent of (ξ, w). See Wiens (1996).

We adopt the viewpoint of *approximate* design theory, and allow as a design measure ξ *any* d.f. on S. It turns out that the optimal designs are then not discrete. It is in fact easy to see that if either of $\sup_f ISB(f;w;\xi)$ or $\sup_g IV(g;w;\xi)$ is to be finite, then ξ must necessarily be absolutely continuous. A formal proof can be based on that of Lemma 1 of Wiens (1992). Our designs may be approximated and implemented by placing the design points at appropriately chosen quantiles of $\xi(x)$.

We say that the pair $(\xi;w)$ is *unbiased* if it satisfies $\sup_f ISB(f;w;\xi) = 0$, and *minimum variance unbiased* (MVU) if it minimizes $\sup_{f,g} IMSE(f;g;w;\xi)$ subject to being unbiased. The following theorem, which is established by standard variational methods, gives a necessary and sufficient condition for unbiasedness, and minimax weights. Before stating it we require some definitions. Let $k(x) = \xi'(x)$ be the design density, and define $m(x) = k(x)w(x)$. Assume, without loss of generality, that the average weight is $\int_S w(x)d\xi(x) = 1$. Then m is a density on S and $\int_S m(x)/w(x)dx = 1$. Define also $l_m(x) = z^T(x)Hz(x)$.

Theorem 1.1 *a) The pair $(\xi;w)$ is unbiased iff $m(x) \equiv \nu$.*

b) For fixed $m(x)$, maximum Integrated Variance is $\sup_g IV(g;w;\xi) = \frac{\sigma^2}{n}\nu^{-1/2}\left(\int_S (w(x)l_m(x)m(x))^2 dx\right)^{1/2}$, attained at the least favourable variance function $g_{m,w}(x) \propto w(x)l_m(x)m(x)$. Maximum IV is minimized by weights $w_m(x) \propto \left(l_m^2(x)m(x)\right)^{-1/3} I(m(x) > 0)$.

c) MVU designs and weights $(\xi_*;w_*)$ for heteroscedastic errors are given by

$$\xi_*'(x) = k_*(x) = \frac{(z(x)^T A^{-1} z(x))^{\frac{2}{3}}}{\int_S (z(x)^T A^{-1} z(x))^{\frac{2}{3}} dx}; \quad w_*(x) = \nu/k_*(x). \qquad (4)$$

The least favourable variances satisfy $g_*(x) = w_*(x)^{-1/2}$. If the errors are homoscedastic ($g \equiv 1$) the exponents $2/3$ in (4) are replaced by $1/2$.

Note that under heteroscedasticity the minimax weights $w_*(x)$ are equal to $g_*(x)^{-2}$; if $g(x)$ is *known* then the *efficient* weights are proportional to $g(x)^{-1}$.

2 SPECIAL CASES

MVU designs for polynomial regression. For $z(x) = (1, x, x^2, \ldots, x^q)^T$ (degree-q polynomial regression) on $S = [-1,1]$, the optimal design densities $k_* = k_*(x;q)$ are most conveniently expressed in terms of orthogonal polynomials. This leads to an interesting connection to the classical

D-optimal design ξ_D, i.e. the discrete measure minimizing the determinant of the covariance matrix of the OLS estimate $\hat{\theta}$. In the following result we denote by $P_m(x)$ the m^{th} degree Legendre polynomial on S, normalized by $\int_{-1}^{1} P_m^2(x)dx = (m+.5)^{-1}$.

Lemma 2.1 *Define a density on S by $h_q(x) = (q+1)^{-1} z^T(x) A^{-1} z(x)$. Then*

$$
\begin{aligned}
k_*(x;q) &\propto h_q(x)^{\frac{2}{3}} = .5(P_q(x)P'_{q+1}(x) - P'_q(x)P_{q+1}(x))^{\frac{2}{3}}; \\
\lim_{q\to\infty} k_*(x;q) &= (1-x^2)^{-\frac{1}{3}}/\{2^{\frac{1}{3}}\beta(2/3,2/3)\}.
\end{aligned}
$$

It can be shown that the local maxima of $h_q(x)$, hence those of $k_*(x;q)$, are the zeros of $(1-x^2)P'_q(x)$. These are precisely the points of support of ξ_D. In this sense, $k_*(\cdot;q)$ is a smoothed version of ξ_D, which has the limiting density $(1-x^2)^{-1/2}/\pi = \lim_{q\to\infty} h_q(x)$. See Figure 1(a) for plots of $k_*(x;3)$ and $k_*(x;\infty)$.

Extrapolation designs. The methods exhibited here apply as well to *extrapolation* designs. Suppose that the goal is the extrapolation of the estimates of the regression response to a region T disjoint from S, with (1) holding on $T \cup S$. Replace the range S of the integrals defining IMSE, ISB and IV by T, obtaining in this way Integrated Mean Squared *Prediction* Error, etc. Redefine H to be $B^{-1}A_T B^{-1}$, where $A_T = \int_T z(x)z^T(x)dx$. Then Theorem 1a),b) apply to the minimization of the resulting IMPSE, subject to $\sup IPB = 0$. The minimax design density under heteroscedasticity is (Fang and Wiens 1997)

$$
k_0(x) = \frac{(z(x)^T A^{-1} A_T A^{-1} z(x))^{\frac{2}{3}}}{\int_S (z(x)^T A^{-1} A_T A^{-1} z(x))^{\frac{2}{3}} dx}
$$

(the exponents are 1/2 for homoscedastic errors) and correspondingly optimal weights are $w_0(x) = Q/k_0(x)$. For degree-q polynomial regression we find that $k_0(x) \propto (\sum_{0\le i,j\le q} \alpha_{ij} P_i(x)P_j(x))^{2/3}$, where $\alpha_{ij} = (i+.5)(j+.5)\int_T P_i(x)P_j(x)dx$. For example, for quadratic regression and a symmetric extrapolation region, i.e. $T = [-\beta, \beta] \setminus S$, we find that $k_0(x) = k_0(x;\beta)$ is given by

$$
k_0(x;\beta) \propto [5\beta^3(\beta+1)(3x^2-1)^2 - \beta(\beta+1)(5x^4-22x^2+5) + 4(1-2x^2+5x^4)]^{\frac{2}{3}};
$$

(an even function of x) while for one-sided extrapolation ($T = [1, \beta]$) we find that

$$
\begin{aligned}
k_0(x;\beta) \propto [&5\beta^4(3x^2-1)^2 + 5\beta^3(3x-1)(x+1)(3x^2-1) - \beta^2(5x^4-30x^3-22x^2 \\
&+10x+5) - \beta(x+1)(5x^3-15x^2-7x+5+2(10x^4+5x^3-4x^2+x+2)]^{\frac{2}{3}}.
\end{aligned}
$$

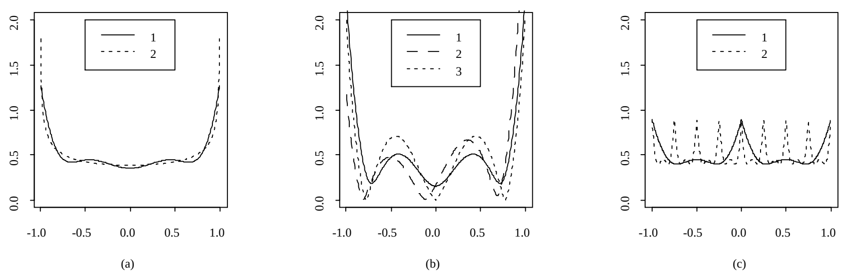

FIGURE 1. MVU design densities. (a) Estimation of polynomial regression. 1: $k_*(x; q = 3)$; 2: $k_*(x; \infty)$. (b) Extrapolation of cubic regression. 1: $k_0(x; \beta = 1.5)$, symmetric extrapolation; 2: $k_0(x; \beta = 1.5)$, asymmetric extrapolation; 3: $k_0(x; \infty)$. (c) Estimation of wavelet regression. 1: $k_{3;0}(x)$; 2: $k_{3;2}(x)$.

For both types of extrapolation region and for fixed q, $k_0(x; \infty) \propto \left(P_q^2(x)\right)^{2/3}$. See Figure 1(b) for plots of $k_0(x; 1.5)$ in both the symmetric and asymmetric cases, and of $k_0(x; \infty)$. For other robust extrapolation designs for polynomial regression see Dette and Wong (1996).

MVU designs for wavelet approximations. Take $S = [0; 1)$ and assume that the regression response $E[Y|x]$ is in the space $L^2(S)$ of square integrable functions, so that it may be approximated arbitrarily closely by linear combinations of multiwavelets (Alpert 1992). Denote by $z_{N;m}(x)$ the $N \cdot 2^{m+1} \times 1$ vector consisting of the wavelets $\{\phi_l(x);\ {}_N\psi_l^{j;k}(x) \mid j = 0; \ldots; m;\ k = 0; \ldots; 2^j - 1;\ l = 0; \ldots; N - 1\}$ in some order. The elements of $z_{N;\infty}$ form an orthogonal basis for $L^2(S)$. When $N = 1$ these coincide with the Haar wavelets, for which design questions were investigated by Herzberg and Traves (1994). The elements of $z_{N;m}(x)$ may be written as $\phi_l(x) = \sqrt{2l+1}P_l(2x - 1)I_{[0;1)}(x)$ and ${}_N\psi_l^{j;k}(x) = 2^{\frac{j}{2}}\,{}_N\psi_l(\{2^j x\})I([2^j x] = k)$; where $\{x\} = x - [x]$ denotes the fractional part of x. The primary wavelets ${}_N\psi_l$ can be developed recursively; examples are ${}_1\psi_0(x) = I_{[0;1/2)}(x) - I_{[1/2;1)}(x)$ and

$$
\begin{aligned}
{}_2\psi_0(x) &= \sqrt{3}(4|x - 1/2| - 1) \cdot I_{[0;1)}(x);\\
{}_2\psi_1(x) &= 2(1 - 3|x - 1/2|) \cdot (I_{[0;1/2)}(x) - I_{[1/2;1)}(x)):
\end{aligned}
$$

By virtue of the orthogonality of $z_{N;m}(x)$, Theorem 1c) can be reduced to a particularly simple form in this case.

Theorem 2.2 (Oyet and Wiens 1997) For the multiwavelet approximation the MVU design density $k_{N;m}(x)$ for homoscedastic errors is

$$
\kappa_N \cdot \left[\phi_{N-1}(\{2^{m+1}x\})\phi'_N(\{2^{m+1}x\}) - \phi'_{N-1}(\{2^{m+1}x\})\phi_N(\{2^{m+1}x\})\right]^{\frac{1}{2}};
$$

where the normalizing constant is

$$\kappa_N = \left(\int_0^1 \left[\phi_{N-1}(x)\phi'_N(x) - \phi'_{N-1}(x)\phi_N(x) \right]^{\frac{1}{2}} dx \right)^{-1} :$$

The MVU weights are $w_{N;m}(x) \propto k_{N;m}(x)^{-1}$.

The exponents $1/2$ are replaced by $2/3$ for heteroscedastic errors. The limiting density is $k_{\infty;m}(x) = \{2^{m+1}x\}^{-1/4}(1 - \{2^{m+1}x\})^{-1/4}/\beta(3/4; 3/4)$: The MVU design can be viewed as a smoothed and dilated version of ξ_D, extended periodically. Some particular cases are $k_{2;m}(x) = 2.51 \cdot [(\{2^{m+1}x\} - 1/2)^2 + 1/12]^{1/2}$ and $k_{3;m}(x) = 8.00 \cdot [((\{2^{m+1}x\} - 1/2)^2 - 1/20]^2 + 1/100)]^{1/2}$. See Figure 1(c) for plots of $k_{3;0}(x)$ and $k_{3;2}(x)$ scaled to $x \in S = [-1; 1)$ for purposes of comparison with the densities in (a) and (b).

REFERENCES

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

Dette, H., and Wong, W.K. (1996), "Robust Optimal Extrapolation Designs," *Biometrika*, 83, 667-680.

Fang, Z., and Wiens, D.P. (1997), "Robust Extrapolation Designs and Weights for Biased Regression Models With Heteroscedastic Errors," *University of Alberta Statistics Centre Technical Report 97.05*.

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

Oyet, A.J., and Wiens, D.P. (1997), "Robust Designs for Wavelet Approximations of Regression Models," *University of Alberta Statistics Centre Technical Report 97.02*.

Wiens, D.P. (1992), "Minimax Designs for Approximately Linear Regression," *Journal of Statistical Planning and Inference*, 31, 353-371.

——(1996), "Asymptotics of Generalized M-Estimation of Regression and Scale With Fixed Carriers, in an Approximately Linear Model," *Statistics and Probability Letters*, 30, 271-285.