Model-Unbiased, Unbiased-in-General Estimation of the Average of a Regression Function

Ken Gerow
Department of Statistics
University of Wyoming
Laramie, WY 82071

Charles E. McCulloch
Biometrics Unit
Cornell University
Ithaca, New York 14853

BU-1229-MA

September 1994

Model-Unbiased, Unbiased-in-General Estimation of the Average of a Regression Function

Ken Gerow and Charles E. McCulloch

Author's Footnote: Ken Gerow is Assistant Professor, Departments of Statistics and Zoology & Physiology, University of Wyoming, Laramie WY 82071. Charles E. McCulloch is Professor, Biometrics Unit and Statistics Centre, Cornell University, Ithaca, NY 14853.
The authors are grateful to the reviewers for their constructive comments.

Abstract

We consider the problem of estimating the mean of a regression function over a finite interval. Classical regression procedures yield conditionally unbiased estimators for that mean (conditioning on the model, and choice of observation points). In contrast, design-based sampling yields estimators that are unconditionally unbiased no matter what the form of the regression model. We propose a class of procedures that enjoy both properties: they are model unbiased, <u>and</u> unbiased-in-general. MSE properties of the class are examined, and illustrative examples are given. The proposed procedures perform well, especially in the typical case where the model is only partially correct.
Key words: model-based, design-based, robust estimation, mean-balanced.

1. INTRODUCTION

Many statistical problems reduce to estimation of an integral over a finite interval, or equivalently, to estimation of the average of some variable over an interval. Examples include block kriging in spatial statistics (Cressie 1991), estimation of average daily metabolic rates (Degen and Kam 1991), bioavailability of various substances in feeding studies (Skelbaek, Anderson, Winning, and Westergaard 1990), toxicological studies (Stone, Spivey Fox and Hogye 1981), and medical studies. In pharmakokinetics, the bioavailability of a drug over time is measured by estimating the area under a curve; the curve represents the instantaneous presence of the drug in the bloodstream (e.g. Bailer and Piegorsch 1990). If the estimate is scaled by the length of the time interval, then the average bioavailability is being estimated.

In this article, we consider the problem of estimating the average of the mean function of a random variable Y over an interval (indexed by X) in which the mean of Y given x is believed to follow a parametric function (with unknown parameters) over the interval.

One approach is to use a parametric regression model: estimate the parameters of the model, and use them directly to estimate the value of the mean. At another extreme of approaches, one can take a simple random sample of observation points, and use the sample mean as an estimator. In the first (model-based) approach, one can derive great benefits if one has chosen the model well; in the second (model-free), one has an unbiased estimator no matter what the true mean function.

We introduce a class of procedures which combine key properties from both of the above approaches. Our estimators are model-unbiased (MU): if the model is correct, the procedure is conditionally unbiased for the average of the function over the interval. Further, they are also unbiased-in-general (UG): the estimators are unconditionally unbiased for the average of an arbitrary mean function. Our proposal is the first broadly applicable class of procedures with both these properties.

The procedures have two components, one in which observation points are deterministically chosen; the other has randomly chosen points. The deterministic component comprises the model-driven element, while the stochastic component attends to robustness. The resulting estimator combines the two components such that the potentially biased model element is corrected to be unbiased.

Until now, MU-UG procedures in a statistical setting have been studied only for polynomial models (Gerow 1984). Our MU-UG procedures are possible for any parametric model. In the sampling literature there are many examples of approximately MU-UG procedures: mean-balanced-sampling (Royall and Cumberland 1981) (which inspired this work), and regression estimators (e.g. Thompson 1992, chap. 8), to name but two. The novelty here is that our procedures are precisely MU and UG.

2. FORMULATION AND PRELIMINARIES

For notational convenience, we consider details of the problem for one dimension. Our model is

$$\begin{aligned} Y_i &= \mu(X_i) + Z_i \text{ ; where} \\ X_i &\in [-1,1] \text{; and} \\ \underset{\sim}{Z} &\sim (0, \sigma^2 \underset{\sim}{I}),\ \underset{\sim}{Z} \text{ independent of } \underset{\sim}{X}. \end{aligned} \qquad (1)$$

In general, $\mu(X)$ need not be parametric. Multiple observations are allowed at a given observation point. The goal is to estimate the average of the mean of Y over [-1,1]:

$$\overline{\mu} = \tfrac{1}{2}\int_{-1}^{1} \mu(t)\,dt, \qquad (2)$$

where μ is assumed to be integrable.

We wish to consider belief in a parametric model $\mu_o(X;\underset{\sim}{\theta})$, linear in $\underset{\sim}{\theta}$. We assume that $\widehat{\underset{\sim}{\theta}} = \widehat{\underset{\sim}{\theta}}(\underset{\sim}{X},\underset{\sim}{Y})$ is unbiased for $\underset{\sim}{\theta}$. We wish to estimate $\overline{\mu}$ subject to constraints which reflect both belief about the regression function and the desire to be robust against misspecification of that function. Correspondingly we define:

Unbiased-in-General. A procedure employing estimator $\widehat{\overline{\mu}} = \widehat{\overline{\mu}}(\underset{\sim}{X}, \underset{\sim}{Y})$ is said to be unbiased in general (UG) for $\overline{\mu}$ if

$$\mathrm{E}_{X,Z}\left(\widehat{\overline{\mu}}\right) = \tfrac{1}{2}\int_{-1}^{1}\mu(x)\,\mathrm{d}x = \overline{\mu} \;\forall \text{ integrable } \mu(x). \tag{3}$$

Model-Unbiased. Given a model $\mu_o(X;\underset{\sim}{\theta})$, a strategy employing estimator $\widehat{\overline{\mu}} = \widehat{\overline{\mu}}(\underset{\sim}{X}, \underset{\sim}{Y})$ shall be called model unbiased (MU) for μ_o if

$$\mathrm{E}_Z\left(\widehat{\overline{\mu}} \mid \underset{\sim}{X} = \underset{\sim}{x}\right) = \tfrac{1}{2}\int_{-1}^{1}\mu_o(x;\underset{\sim}{\theta})\,\mathrm{d}x. \tag{4}$$

Strategies that satisfy (3) and (4) will be labeled MU-UG strategies; that the MU part of the label depends upon a choice of model will be suppressed, but understood. MU and UG reflect the model input and desire for robustness in the sampling situation.

3. THE PROPOSED MU-UG STRATEGY

The basic idea behind the proposed strategy is to divide the sample into two parts. One part is used to satisfy the model constraint; the other to satisfy the robustness constraint. To the model (1), we add the following details.

$$\begin{aligned} \underset{\sim}{X}' &= [\underset{\sim}{X}'_m \;\vdots\; \underset{\sim}{X}'_r] \text{ , where} \\ &\underset{\sim}{X}_m \text{ is fixed, } \underset{\sim}{X}_r \sim f(\underset{\sim}{x}_r)\text{, and} \\ n &= n_m + n_r \text{ is the sample size.} \end{aligned} \tag{5}$$

Subscripts m and r refer to the model and robustness components of the procedure, respectively. $\underset{\sim}{Y}$ is accordingly partitioned into $\underset{\sim}{Y}_m$ and $\underset{\sim}{Y}_r$, similarly for $\underset{\sim}{Z}$.

Model input to the procedure is as follows. Using $\widehat{\underset{\sim}{\theta}}_m = \widehat{\underset{\sim}{\theta}}(\underset{\sim}{X}_m, \underset{\sim}{Y}_m)$, an estimator for the mean of μ_o is

$$\widehat{\overline{\mu}}_{o,m} = \tfrac{1}{2}\int_{-1}^{1}\mu_o(x;\widehat{\underset{\sim}{\theta}}_m)\,\mathrm{d}x, \tag{6}$$

which is unbiased for $\overline{\mu}$ if $\mu = \mu_o$.

The robustness input to the procedure is to apply *any* UG procedure for $\overline{\mu}$ to $(\underset{\sim}{Y}_r, \underset{\sim}{X}_r)$. That one always exists is assured: a simple random sample (SRS) of observation points $\underset{\sim}{X}_r$, with $\widehat{\widehat{\mu}}=\overline{Y}_r$ is UG. For simplicity and later to derive an unbiased estimate of $\mathrm{Var}(\widehat{\widehat{\mu}})$, we will use this strategy herein. The pair $\left(\overline{Y}_r, f(\underset{\sim}{x}_r)\right)$, where f is the distribution used to generate $\underset{\sim}{X}_r$, we call the UG kernel of the associated MU-UG strategy.

Our procedure has as its estimator

$$\widehat{\widehat{\mu}} \;=\; \widehat{\widehat{\mu}}_{o,m} \;+\; \overline{Y}_r \;-\; \overline{\mu}_{o,r} \;, \tag{7}$$

where

$$\overline{\mu}_{o,r} \;=\; \tfrac{1}{r}\sum_{i=1}^{r}\mu_o(x_{i,r};\widehat{\underset{\sim}{\theta}}_m);$$

in words, $\overline{\mu}_{o,r}$ is the average of values fitted to the random points, but with estimation of $\underset{\sim}{\theta}$ coming from the model points. One way to think about this estimator is to note that the first term, $\widehat{\widehat{\mu}}_{o,m}$, is a model-based estimator of $\overline{\mu}$, which may be biased if μ_o does not hold. The remainder of (7) corrects this potential bias. We proceed with an example, followed by a proof of the MU-UG property of the procedure.

<u>Example 1</u>. Suppose $\mu_o(x;\underset{\sim}{\theta})=\theta_1+\theta_2 x$, $\underset{\sim}{\theta}\in\mathbb{R}^2$, is the belief model. We use the UG kernel $\left(\overline{Y}_r,\ \underset{\sim}{X}_r\sim \mathrm{SRS}\right)$. Let $\mu_o(x;\widehat{\underset{\sim}{\theta}})$ be determined by least squares, based on the model observations. Thus the first term of (7) is, by (6), $\widehat{\widehat{\mu}}_{o,m}=\widehat{\theta}_1$. Then,

$$\widehat{\widehat{\mu}} \;=\; \widehat{\theta}_1 \;+\; \overline{Y}_r \;-\; \tfrac{1}{r}\sum_{i=1}^{r}\left(\widehat{\theta}_1+\widehat{\theta}_2 x_{i,r}\right). \tag{8}$$

To see that (8) is MU, note that if μ_o obtains,

$$\begin{aligned}
\mathrm{E}\left(\widehat{\widehat{\mu}}\mid \underset{\sim}{x}\right) &= \theta_1 + \tfrac{1}{r}\sum_{i=1}^{r}\left(\theta_1+\theta_2 x_{i,r}\right) - \tfrac{1}{r}\sum_{i=1}^{r}\left(\theta_1+\theta_2 x_{i,r}\right)\;, \\
&\qquad \text{(since } \mathrm{E}\left(\widehat{\underset{\sim}{\theta}}\mid \underset{\sim}{x}\right)=\underset{\sim}{\theta},\ \text{and } \mathrm{E}(\underset{\sim}{Z}\mid \underset{\sim}{x})=\underset{\sim}{0}) \\
&= \theta_1 \;=\; \overline{\mu}_o .
\end{aligned}$$

As for UG, note that $\mathrm{E}_{X,Z}\left(\overline{Y}_r\right)=\overline{\mu}$ for any μ. Since $\underset{\sim}{Z}_m$ is independent of $\underset{\sim}{X}_r$, the last term of (8) has

$$\begin{aligned}
\mathrm{E}_{X,Z}\left(\frac{1}{r}\sum_{i=1}^{r}\left(\widehat{\theta}_1+\widehat{\theta}_2 X_{i,r}\right)\right) &= \mathrm{E}_{Z_m\mid x}\left(\widehat{\theta}_1\right) + \mathrm{E}_{Z_m\mid x}\left(\widehat{\theta}_2\right)\mathrm{E}_{X_r}\left(X_{1,r}\right) \\
&= \mathrm{E}_{Z_m\mid x}\left(\widehat{\theta}_1\right) \text{ (since } \mathrm{E}_{X_r}\left(X_{1,r}\right) \;=\; 0).
\end{aligned}$$

This expectation is a function of the choice of $\underset{\sim}{x}_m$ and the unspecified μ. Note that it is also the expectation of the last term of (8); thus

$$\begin{aligned} \mathrm{E}_{X,Z}\left(\widehat{\overline{\mu}}\right) &= \mathrm{E}_{Z_m|X}\left(\widehat{\theta}_1\right) + \mathrm{E}_{X,Z}\left(\overline{Y}_r\right) - \mathrm{E}_{Z_m|X}\left(\widehat{\theta}_1\right) \\ &= \mathrm{E}_{X,Z}\left(\overline{Y}_r\right) = \overline{\mu}. \end{aligned}$$

For the simple linear belief model of this example, the estimator looks like a classical difference estimator, or, for that matter, a regression estimator (e.g. Thompson 1992). The choice of virtually any other belief model dispells that apparent identity.

It is convenient for what follows to write the true regression function in (1) as

$$\mu(X) = \underset{\sim}{h}(X)'\underset{\sim}{\theta} + \gamma(X) \quad (\underset{\sim}{\theta} \text{ is length } p). \tag{9}$$

Thus $\overline{\mu} = \frac{1}{2}\underset{\sim}{H}'\underset{\sim}{\theta} + \overline{\gamma}$, where $\underset{\sim}{H} = \left\{\int_{-1}^{1} h_i(x)\mathrm{d}x\right\}_p$, and $\overline{\gamma} = \frac{1}{2}\int_{-1}^{1}\gamma(x)\,\mathrm{d}x$. If $\mu = \mu_o$, then $\gamma \equiv 0$. If $\mu \neq \mu_o$, $\underset{\sim}{\theta}$ is defined conditional upon the design points and $\underset{\sim}{\gamma}_m$. In particular, if we denote the least squares estimate of $\underset{\sim}{\theta}$ by $\underset{\sim}{D}'\underset{\sim}{Y}$, then $\underset{\sim}{D}'\underset{\sim}{\gamma}_m = 0$.

Theorem 3.1. MU-UG Theorem for the proposed strategy.
We use model (1), with details of $\mu(X)$ given in (9) and the following assumptions:

(A1) $\widehat{\underset{\sim}{\theta}}$ is the least squares estimator for $\underset{\sim}{\theta}$;
(A2) $\left(\overline{Y}_r, f(x_r)\right)$ is UG for $\overline{\mu}$; and

If (A1) and (A2) obtain, the estimator (7) is MU under μ_o, and UG.
Proof: See the appendix.
Remarks:
1. If μ_o is nonlinear in $\underset{\sim}{\theta}$ and $\widehat{\underset{\sim}{\theta}}$ is asymptotically unbiased, then the procedure is UG and asymptotically MU. Proof follows that for Theorem 3.1, employing the usual asymptotic tools (e.g. Seber and Wild 1989).
2. If function evaluations are made without error (a numerical quadrature setting), then for both linear <u>and</u> nonlinear models,

least-squares $\hat{\underset{\sim}{\theta}}$ are exact for $\underset{\sim}{\theta}$. Then the strategy yields exact answers under μ_o, and is UG.

3. The procedure can easily be extended to higher dimensional $\underset{\sim}{X}$. All we need is a model for which $\hat{\underset{\sim}{\theta}}$ is unbiased; the strategy ($\overline{Y}_r$, $\underset{\sim}{X}_r \sim$ SRS) is still UG.

3.2. Variance of the Proposed MU-UG Estimation Strategies.

We now develop the variance of the procedure, for which we need some notational apparati:

N1. Write $r\sigma^2_{\gamma,r} := \mathrm{E}_X\left(\underset{\sim}{1}'_r(\underset{\sim}{\gamma}_r - \overline{\gamma})\right)^2 = \int_{-1}^{1}\cdots\int_{-1}^{1}\left(\underset{\sim}{1}'_r(\underset{\sim}{\gamma}_r - \overline{\gamma})\right)^2 f(\underset{\sim}{x}_r)\mathrm{d}\underset{\sim}{x}$, where $f(\underset{\sim}{x}_r)$ is the distribution of the "robustness points", $\underset{\sim}{\gamma}_r$ is $\gamma(\underset{\sim}{X}_r)$, and $\mathrm{E}_{X,Z}(\underset{\sim}{1}'\underset{\sim}{\gamma}_r) = r\overline{\gamma}$. Note that under simple random sampling,

$$\sigma^2_{\gamma,r} = \tfrac{1}{2}\int_{-1}^{1}\left(\gamma(x) - \overline{\gamma}\right)^2 \mathrm{d}x.$$

N2. The design component has n_m observations (chosen so that n_m/p is an integer); we write $\underset{\sim}{M}$ as an (n_m/p)-replicate of the (assumed to exist) minimal design $\underset{\sim}{M}_*$:

$$\underset{\sim}{M}_* = \left\{h_j(x_{i,m})\right\}_{p\times p}.$$

N3. By (A2), $(\overline{Y},f)$ is UG so $\mathrm{E}_{X,Z}\left(\underset{\sim}{1}'_r\underset{\sim}{R}\right) = \frac{r}{2}\underset{\sim}{H}$; write $\underset{\sim}{V} = \mathrm{Var}\left(\underset{\sim}{1}'_r\underset{\sim}{R}\right)$, where $\underset{\sim}{R}$ is the observation matrix $\left\{h_j(x_{i,r})\right\}_{n_r,p}$.

Theorem 3.2. Under the assumptions of Theorem 3.1, for linear μ_o,

$$\mathrm{Var}\left(\hat{\overline{\mu}}\right) = \frac{1}{n_r}\left\{\sigma^2_{\gamma,r} + \sigma^2_z + \frac{p}{n_r n_m}\sigma^2_z \mathrm{tr}\left((\underset{\sim}{M}'_*\underset{\sim}{M}_*)^{-1}\underset{\sim}{V}\right)\right\}. \tag{10}$$

Proof: See the appendix.

Remarks:

1. If the model is correct, $\sigma^2_{\gamma,r} = 0$ in (10).
2. If the UG kernel is ($\overline{Y}_r$, $\underset{\sim}{X}_r \sim$ SRS), then

$$\mathrm{Var}\left(\hat{\overline{\mu}}\right) = \frac{1}{n_r}\left\{\sigma^2_{\gamma,r} + \sigma^2_z + \frac{p}{n_m}\sigma^2_z \mathrm{tr}\left((\underset{\sim}{M}'_*\underset{\sim}{M}_*)^{-1}\underset{\sim}{V}_1\right)\right\}, \tag{11}$$

where $\underset{\sim}{V}_1$ is $\underset{\sim}{V}$ for $n_r = 1$. The variance in this form is particularly amenable to deducing optimal allocation of n_m and n_r.

3. In an application where belief in μ_o is strong, it may be reasonable to assume that $\sigma^2_{\gamma,r} < \sigma^2_z$. Then

$$\text{Var}\left(\hat{\bar{\mu}}\right) \leq \frac{\sigma^2_z}{n_r}\left\{2 + \frac{p}{n_r n_m}\text{tr}\left(\left(\underset{\sim}{\mathbf{M}}'_*\underset{\sim}{\mathbf{M}}_*\right)^{-1}\underset{\sim}{\mathbf{V}}\right)\right\}. \tag{12}$$

4. The variance under nonlinear models is analogous to the linear one; asymptotically it is

$$\text{Var}\left(\hat{\bar{\mu}}\right) \approx \frac{1}{n_r}\left(\sigma^2_{\gamma,r} + \sigma^2_z + \frac{p}{n_r n_m}\sigma^2_z\text{tr}\left(\left(\underset{\sim}{\mathbf{U}}'_*\underset{\sim}{\mathbf{U}}_*\right)^{-1}\underset{\sim}{\mathbf{V}}\right)\right),$$

where

$$\underset{\sim}{\mathbf{U}}_{m\times p} = \underset{\sim}{\mathbf{U}}(\underset{\sim}{\boldsymbol{\theta}}) = \left(\frac{\partial\mu_o(x_i;\theta)}{\partial\theta_j}\right) \; ; \; i = 1,2, \cdots n_m;$$

$j = 1, 2, \cdots, p$, and where $\underset{\sim}{\mathbf{V}}$ is the variance-covariance matrix of $\underset{\sim}{a}$:

$$\underset{\sim}{a}' = \frac{1}{n_r}\left(\sum\partial\mu_o(x_i;\underset{\sim}{\theta})/\partial\theta_1, \sum\partial\mu_o(x_i;\underset{\sim}{\theta})/\partial\theta_2, \cdots, \sum\partial\mu_o(x_i;\underset{\sim}{\theta})/\partial\theta_p\right).$$

Note: $\underset{\sim}{\mathbf{U}}$ is a function of $\underset{\sim}{x}_m$ only, so is independent of $\underset{\sim}{X}_r$.

When the UG kernel is $(\bar{Y}_r, \underset{\sim}{X}_r \sim \text{SRS})$, we have a surprisingly simple estimator of the variance (10). Let $\underset{\sim}{\delta} = \underset{\sim}{Y}_r - \underset{\sim}{\hat{Y}}_r$ be the residuals at the random observation points, and s^2_δ be the sample variance among them. An estimator for $\text{Var}\left(\hat{\bar{\mu}}\right)$ is as follows.

Theorem 3.3. Unbiased Estimator of $\text{Var}\left(\hat{\bar{\mu}}\right)$.

Under the conditions of Theorem 3.1, and with the UG kernel $(\bar{Y}_r, \underset{\sim}{X}_r \sim \text{SRS})$,

$$\text{E}_{X,Z}\left(\frac{s^2_\delta}{n_r}\right) = \frac{1}{n_r}\left\{\sigma^2_{\gamma,r} + \sigma^2_z + \frac{p}{n_m}\sigma^2_z\text{tr}\left(\left(\underset{\sim}{\mathbf{M}}'_*\underset{\sim}{\mathbf{M}}_*\right)^{-1}\underset{\sim}{\mathbf{V}}_1\right)\right\}$$
$$= \text{Var}\left(\hat{\bar{\mu}}\right).$$

Proof: See the appendix.

Remark: If one has multiple observations at the model points with which to estimate "pure error" (σ^2_z), one can decompose the variance estimator to arrive at an estimate of σ^2_γ.

4. EXAMPLES AND COMPARISONS

The variance (10) suggests that the optimal design for $\underset{\sim}{x}_m$ is given by A-optimality, i.e. by minimizing the trace of $(\underset{\sim}{M}'\underset{\sim}{M})^{-1}$. Given a choice of design $\underset{\sim}{x}_m$, we can then use (11) to optimally choose m and r, given n. In particular, if one assumes that the model is nearly correct, i.e. assume $\sigma_\gamma^2 \approx 0$, the calculations are generally quite tractable. To facilitate the examples, we will assume that $\underset{\sim}{X}_r$ is chosen via a SRS, so that we can apply the simplifying features of (12) and Theorem 3.3. The following example demonstrates these aspects.

<u>Example 1, continued.</u> Suppose $\mu_o(x;\underset{\sim}{\theta}) = \theta_1 + \theta_2 x$. We will consider two cases: the model is correct, and that the true function is in fact $\mu(x;\theta) = \theta_1 + \theta_2 x + \theta_3 x^2$. We will compare our strategy to a simple random sample and to a purely model based strategy: take all n observations at $x = 0$, and use $\widehat{\widehat{\mu}} = \overline{Y}$.

The correct A-optimal design for $\underset{\sim}{x}_m$ is to have an equal number of observations at 1 and -1. Under the model, and using simple random sampling for choosing $\underset{\sim}{X}_r$, (11) reduces to

$$\text{Var}(\widehat{\widehat{\mu}}) \le \frac{1}{n_r}\left(1 + \frac{1}{3n_m}\right).$$

From this, we deduce that the optimal sample apportionment is $n_m = 2$, for $n \le 30$; it becomes $n_m = 4$ for $31 \le n \le 82$, and jumps to $n_m = 6$ at that point.

The MSE can be decomposed into functions of parameters only and of σ_z^2. The MSE components for this example are displayed in Table 1. We can examine this MSE behavior visually for any set of parameter values. For example, to illustrate a case where the model is incorrect, set $\theta_2 = 2$, $\theta_3 = 2$. Figure 1 shows the ratios $\frac{\text{Var(MU-UG)}}{\text{Var(SRS)}}$, and $\frac{\text{Var(MU-UG)}}{\text{Var(model)}}$, for a range of n.

The behavior we see here is a general property of the MU-UG strategy. If the model is correct, one can do somewhat better by wedding oneself to the model. If the model is incorrect, one can (depending on the confluence of parameter values and σ_z^2), often experience significant gains by having "partially" specified the true mean function, while limiting the damage that could be experienced by having wedded oneself to the incorrect model. Note that if θ_3 is very small, the MU-UG strategy will not necessarily be competitive against the pure model-based strategy. If there is no regression ($\mu = \theta_1$), then the SRS strategy also outperforms the MU-UG strategy.

Example 2. We illustrate the use of our MU-UG technique on a subset of a well-known spatial statistics data set, the Wolfcamp-aquifer data (Cressie 1991, p. 212). Cressie gives the locations (x_1, x_2), and the piezometric-head values, Y, for 85 wells located near Amarillo, Texas. We take as our inferential goal the estimation of the average of the piezometric-head value over the rectangle given by $-45 \leq x_1 \leq 105$ and $10 \leq x_2 \leq 210$.

Since our procedure has a deterministic and a stochastic component we will pretend that the observations at the (x_1, x_2) values given by (-2.23054,29.91113), (103.26625,20.34329), (42.78275,127.62282), (83.14496,159.11558), and (-24.06744, 184.76636) are the model-based, deterministic observations and that the remainder of the observations with $-45 \leq x_1 \leq 105$ and $10 \leq x_2 \leq 210$ represent a random, uniformly distributed sample from the rectangle. So we have 5 deterministic and 64 stochastic observations. We could also arrive at this situation by starting with a SRS of points and operating conditionally on a randomly chosen subset.

As a simple belief model which a geologist might have had before the data were gathered, we take $\mu_{o,i} = \theta_0 + \theta_1 x_{1i} + \theta_2 x_{2i}$. Using the model observations gives $\hat{\theta}_0$ = 2502.4817, $\hat{\theta}_1$ = -5.9218, and $\hat{\theta}_2$ = -5.4813, so that

$$\overline{\mu}_{o,r} = \frac{1}{r}\sum_{i=1}^{r}\left(\hat{\theta}_0 + \hat{\theta}_1 x_{1i,r} + \hat{\theta}_2 x_{2i,r}\right)$$

$$= \hat{\theta}_0 + \hat{\theta}_1 \overline{x}_{1\cdot,r} + \hat{\theta}_2 \overline{x}_{2\cdot,r} = 1838.8886;$$

$$\overline{\mu}_{o,m} = \frac{1}{(210-10)(105+45)} \int_{10}^{210} \int_{-45}^{105} \left(\hat{\theta}_0 + \hat{\theta}_1 x_1 + \hat{\theta}_2 x_2 \right) dx_1 dx_2$$

$$= \hat{\theta}_0 + \hat{\theta}_1 30 + \hat{\theta}_2 110 = 1721.5556 \text{ ; and}$$

$$\overline{y}_r = 1844.5556.$$

We therefore have

$$\hat{\hat{\mu}} = 1721.5556 + 1844.5556 - 1838.8886 = 1727.5544,$$

with a standard error of $s_\delta/\sqrt{n} = 24.6516$.

For comparison, treating the entire sample as a SRS gives

$$\hat{\hat{\mu}}_{SRS} = \overline{y}_{SRS} = 1833.2059,$$

and using a strictly model-based approach gives

$$\hat{\hat{\mu}}_{MOD} = \hat{\theta}_0 + \hat{\theta}_1 30 + \hat{\theta}_2 110$$

$$= 2589.3885 - 6.8993(30) - 6.0387(110)$$

$$= 1718.1417,$$

where the $\hat{\theta}_i$ are now calculated from the entire data set. In this case, the MU-UG approach yields an answer intermediate to the other two techniques.

5. DISCUSSION

We considered the problem of estimating the mean of a regression function over an interval, where some parametric model is supposed but not assumed for the function. We have presented a class of procedures for that problem, members of which have the property of being simultaneously MU and UG. The strategy employs explicit model-based and robustness-based components, combining the two in a special way.

The variance of the strategy was obtained in general, as well as an unbiased estimator for it in the special case where the random points are chosen according to a SRS. Examination of the variance of the proposed strategy, both in its general form, and through an example, reveals that it works as one might expect. If the model is only partially correct (a realistic situation), one can gain greatly using our strategy over a model-dedicated approach, while still being better than a simple random sample.

The strategy can be applied even if some of the independent variables can have some known non-uniform distribution over the range for which one wishes to estimate the mean. The concept of a SRS over the interval (which induces a uniform distribution) simply has to be modified to reflect the joint distribution of the variables of interest.

APPENDIX: PROOFS

Proof of Theorem 3.1. Additional Notation: Let

$$\underset{\sim}{1}_r = \left\{1,1,\cdots,1\right\}' \text{ be of length } r;$$

$$\mu(x) = \underset{\sim}{h}(x)'\underset{\sim}{\theta} + \gamma(x); \ \underset{\sim}{h} \text{ and } \underset{\sim}{\theta} \text{ are of length } p;$$

$$\underset{\sim}{H} = \left\{\underset{\sim}{H}_i\right\}_p = \left\{\int_{-1}^{1} h_i(x)\mathrm{d}x\right\}_p$$

$$\underset{\sim}{\mathbf{M}}_{(m\times p)} = \left\{h_j(x_i)\right\}; i=1,\cdots,n_m \ (\underset{\sim}{\mathbf{M}} \text{ is fixed through } \underset{\sim}{x}_m);$$

$$\underset{\sim}{R}_{(r\times p)} = \left\{h_j(X_i)\right\}; \ i=n_m+1,\cdots,n \ (\underset{\sim}{R} \text{ is stochastic through } \underset{\sim}{X}_r);$$

$$\underset{\sim}{Y}_m = \underset{\sim}{\mathbf{M}}\underset{\sim}{\theta} + \underset{\sim}{\gamma}_m + \underset{\sim}{Z}_m \quad (\text{analogously for } \underset{\sim}{Y}_r);$$

$$\underset{\sim}{\mathbf{D}}' = (\underset{\sim}{\mathbf{M}}'\underset{\sim}{\mathbf{M}})^{-1}\underset{\sim}{\mathbf{M}}', \text{ so } \underset{\sim}{\widehat{\theta}}_m = \underset{\sim}{\mathbf{D}}'\underset{\sim}{Y}_m; \text{ and}$$

$$\underset{\sim}{\widehat{\mu}}_{o,r} = \underset{\sim}{h}(\underset{\sim}{X}_r)'\underset{\sim}{\widehat{\theta}} \quad (\text{estimates of } \mu_o \text{ at } \underset{\sim}{X}_r).$$

With this machinery in hand,

$$\begin{aligned}\widehat{\widehat{\mu}} &= \tfrac{1}{r}\underset{\sim}{1}_r'\underset{\sim}{Y}_r - \tfrac{1}{r}\underset{\sim}{1}_r'(\underset{\sim}{\widehat{\mu}}_{o,r}) + \tfrac{1}{2}\underset{\sim}{H}'\underset{\sim}{\widehat{\theta}} \\ &= \tfrac{1}{r}\underset{\sim}{1}_r'\left(\underset{\sim}{R}\underset{\sim}{\theta} + \underset{\sim}{\gamma}_r + \underset{\sim}{Z}_r - \underset{\sim}{R}\underset{\sim}{\theta} - \underset{\sim}{R}\underset{\sim}{D}'\underset{\sim}{Z}_m\right) + \tfrac{1}{2}\underset{\sim}{H}'\left(\underset{\sim}{\theta} + \underset{\sim}{D}'\underset{\sim}{Z}_m\right) \\ &= \tfrac{1}{r}\underset{\sim}{1}_r'\left(\underset{\sim}{\gamma}_r + \underset{\sim}{Z}_r - \underset{\sim}{R}\underset{\sim}{D}'\underset{\sim}{Z}_m\right) + \tfrac{1}{2}\underset{\sim}{H}'\left(\underset{\sim}{\theta} + \underset{\sim}{D}'\underset{\sim}{Z}_m\right). \qquad (A.1)\end{aligned}$$

Thus $\mathrm{E}_{Z|x}(\widehat{\widehat{\mu}}) = \tfrac{1}{r}\underset{\sim}{1}_r'\underset{\sim}{\gamma}_r + \tfrac{1}{2}\underset{\sim}{H}'\underset{\sim}{\theta}$.

If the model is true, $\gamma(x) \equiv 0$ by definition: the estimator is MU. Otherwise, $\mathrm{E}_{Z,X}(\widehat{\widehat{\mu}}) = \mathrm{E}\left(\mathrm{E}_{Z|x}(\widehat{\widehat{\mu}})\right) = \tfrac{1}{r}\underset{\sim}{1}_r'\underset{\sim}{1}_r\overline{\gamma} + \tfrac{1}{2}\underset{\sim}{H}'\underset{\sim}{\theta} = \overline{\mu}$. □

Proof of Theorem 3.2. From (A.1) and using $\overline{\mu} = \tfrac{1}{2}\underset{\sim}{H}'\underset{\sim}{\theta} + \overline{\gamma}$,

$$\begin{aligned}\mathrm{Var}\left(\widehat{\widehat{\mu}}\right) &= \mathrm{E}\left[\left(\widehat{\widehat{\mu}} - \overline{\mu}\right)^2\right] \\ &= \left\{\left(\tfrac{1}{n_r}\underset{\sim}{1}_r'\underset{\sim}{\gamma}_r - \overline{\gamma}\right) + \tfrac{1}{n_r}\underset{\sim}{1}_r'\underset{\sim}{Z}_r + \left(\tfrac{1}{2}\underset{\sim}{H}' - \tfrac{1}{n_r}\underset{\sim}{1}_r'\underset{\sim}{R}\right)\underset{\sim}{\mathbf{D}}'\underset{\sim}{Z}_m\right\}^2\end{aligned}$$

$$= \text{Var}\left(\tfrac{1}{n_r}\underset{\sim}{1}'_r\underset{\sim}{\gamma}_r - \bar{\gamma}\right) + \text{Var}\left(\tfrac{1}{n_r}\underset{\sim}{1}'_r\underset{\sim}{Z}_r\right) + \text{Var}\left(\left(\tfrac{1}{2}\underset{\sim}{H}' - \tfrac{1}{n_r}\underset{\sim}{1}'_r\underset{\sim}{R}\right)\underset{\sim}{D}'\underset{\sim}{Z}_m\right)$$

(since all cross-product terms have expectation zero)

$$= \frac{1}{n_r^2}\left\{ n_r\sigma^2_{\gamma,r} + n_r\sigma^2_z + \sigma^2_z \text{tr}[\underset{\sim}{D}'\underset{\sim}{D}\underset{\sim}{V}] \right\},$$

This last equality holding by applying Theorem 1, p. 55 of Searle (1971) twice. Since $\underset{\sim}{D}'\underset{\sim}{D} = \frac{p}{n_m}\left(\mathbf{M}_*'\mathbf{M}_*\right)^{-1}$,

$$\text{Var}\left(\hat{\bar{\mu}}\right) = \frac{1}{n_r}\left\{\sigma^2_{\gamma,r} + \sigma^2_z\left(1 + \frac{p}{n_r n_m}\text{tr}\left(\left(\underset{\sim}{M}_*'\underset{\sim}{M}_*\right)^{-1}\underset{\sim}{V}\right)\right)\right\}. \quad \square$$

Proof of Theorem 3.3. Since the δ_i are exchangeable, they have the same means and variances and are equicorrelated. It is therefore easy to show that $E\left(s^2_\delta\right) = \text{Var}(\delta_i) - \text{Cov}(\delta_i, \delta_j)$. It remains to calculate the variance and covariance. The variance is

$$E_{X,Z}\left(\delta_i - \mu_\delta\right)^2 = E_{X,Z}\left\{ Y_i - \hat{Y}_i - \bar{\gamma}\right\}^2$$

$$= E_{X,Z}\left\{\underset{\sim}{h}'(X_i)\underset{\sim}{\theta} + \gamma(X_i) + Z(X_i) - \underset{\sim}{h}'(X_i)\hat{\underset{\sim}{\theta}} - \bar{\gamma}\right\}^2$$

$$= E_{X,Z}\left\{ \left(\gamma(X_i) - \bar{\gamma}\right) + Z(X_i) - \underset{\sim}{h}'(X_i)\underset{\sim}{D}'\underset{\sim}{Z}_m \right\}^2$$

$$= \left\{\sigma^2_{\gamma,r} + \sigma^2_z\left(1 + \frac{p}{n_r n_m}\text{tr}\left(\left(\underset{\sim}{M}_*'\underset{\sim}{M}_*\right)^{-1}\underset{\sim}{V}\right) + \sigma^2_z \mu_h'\underset{\sim}{D}'\underset{\sim}{D}\mu_h\right)\right\}. \tag{A.2}$$

The crossproduct expectation is

$$E_{X,Z}\left(\delta_i\delta_j\right) = E_{X,Z}\left\{ (Y_i - \hat{Y}_i)(Y_j - \hat{Y}_j)\right\}$$

$$= E_{X,Z}\left\{\left(\gamma(X_i) + Z(X_i) - \underset{\sim}{h}'(X_i)\underset{\sim}{D}'\underset{\sim}{Z}_m\right)\left(\gamma(X_j) + Z(X_j) - \underset{\sim}{h}'(X_j)\underset{\sim}{D}'\underset{\sim}{Z}_m\right)\right\}$$

(the terms $Z(X_i)$, $Z(X_j)$, and $\underset{\sim}{Z}_m$ have expectation zero and are mutually independent, so all their associated crossproduct terms can be conveniently ignored)

$$= E_{X,Z}\left\{\gamma(X_i)\gamma(X_j) + \underset{\sim}{h}'(X_i)\underset{\sim}{D}'\underset{\sim}{Z}_m\underset{\sim}{Z}'_m\underset{\sim}{D}\underset{\sim}{h}(X_j)\right\}$$

(to the last term, apply the bilinear version of Theorem 1, p. 55, Searle (1971), denoting $E\left(\underset{\sim}{h}'(X_i)\right)$ by μ_h)

$$= \bar{\gamma}^2 + \sigma_z^2 \mu_h' \underset{\sim}{D}' \underset{\sim}{D} \mu_h .$$

Thus the covariance is $\sigma_z^2 \mu_h' \underset{\sim}{D}' \underset{\sim}{D} \mu_h$. Subtract this term from (A.2), divide by n_r, and the result is proved. □

REFERENCES

Bailer, A. J, and Piegorsch, W. W. (1990), "Estimating Integrals Using Quadrature Methods With an Application in Pharmakokinetics," Biometrics 46(4):1201-1211.

Cressie, N.A.C. (1991), Statistics for Spatial Data, New York: John Wiley.

Degen, A. A. and Kam, M. (1991), "Average Daily Metabolic Rate of Gerbils of Two Species: Gerbillus pyramidum and Gerbillus allenbyi," Journal of Zoology, 223(1), 143-149.

Gerow, K. (1984), An Algorithm for Random Balanced Sampling, M.Sc. thesis, University of Guelph, Guelph, Ontario, Canada.

Royall, R.M., and Cumberland, W.G. (1981), "An Empirical Study of the Ratio Estimator and Estimators of its Variance," Journal of the American Statistical Association 76,66-77.

Searle, S.R. (1971), "Linear Models," New York: John Wiley.

Seber, G. A. F. and Wild, C.J. (1989), Nonlinear Regression, New York: John Wiley.

Skelbaek, T., Andersen, N.G., Winning, M., and Westergaard, S. (1990), "Stability in Fish Feed and Bioavailability to Rainbow Trout of Two Ascorbic Acid Forms," Aquaculture, 84(3), 335-343.

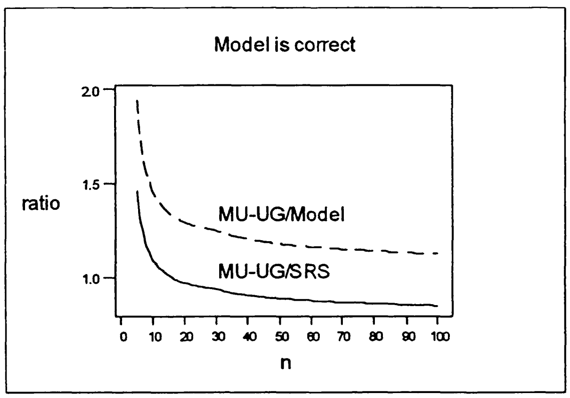
Stone, C.L., Spivey Fox, M.R., and Hogye, K.S. (1981), "Bioavailability of Lead in Oysters Fed to Young Japanese Quail," Environmental Research, 26(2), 409-421.

Thompson, S.K. (1992), Sampling, New York: John Wiley.

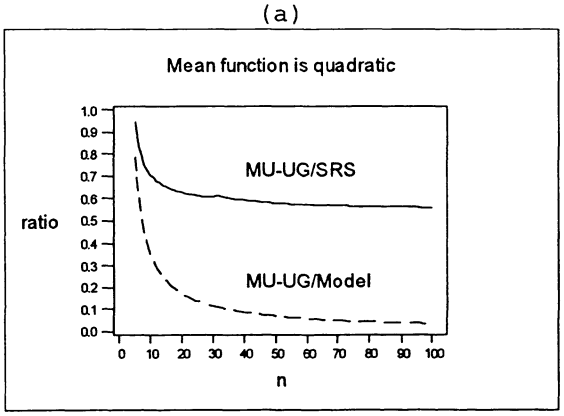
Table 1. MSE components of estimators under an assumed model $\mu_o=\theta_1+\theta_2x$ and also when $\mu=\theta_1+\theta_2x+\theta_3x^2$. Each MSE has a σ_z^2 (error) component and a model component.

	MSE Under the Model	
	$\mu=\theta_1+\theta_2x$	$\mu=\theta_1+\theta_2x-\theta_3x^2$
MU-UG	$\frac{\sigma_z^2}{n_r-n_m}\left(1+\frac{1}{3n_m}\right)+0$	$\frac{\sigma_z^2}{n_r-n_m}\left(1+\frac{1}{3n_m}\right)+\frac{4\theta_3^2}{45(n_r-n_m)}$
SRS	$\frac{\sigma_z^2}{n}+\frac{\theta_2^2}{3n}$	$\frac{\sigma_z^2}{n}+\frac{1}{n}\left(\frac{\theta_2^2}{3}+\frac{4\theta_3^2}{45}\right)$
model-based	$\frac{\sigma_z^2}{n}+0$	$\frac{\sigma_z^2}{n}+\frac{\theta_3^2}{9}$

Figure 1. Plot of $\frac{\text{Var(MU-UG)}}{\text{Var(SRS)}}$, and $\frac{\text{Var(MU-UG)}}{\text{Var(model)}}$ for $\mu_1=\theta_1+2x$ (a) and $\mu_2=\theta_1+2x+2x^2$ (b), with $\mu_o=\theta_1+\theta_2x$; $\sigma_z^2=1$.



(a)



(b)