IMPROVED ESTIMATORS OF MEAN RESPONSE IN SIMULATION WHEN CONTROL VARIATES ARE USED

by

Leon Jay Gleser
Purdue University

Technical Report #87-32

Department of Statistics
Purdue University

July 1987

★

IMPROVED ESTIMATORS OF MEAN RESPONSE IN SIMULATION WHEN CONTROL VARIATES ARE USED

Leon Jay Gleser[1]
Purdue University

ABSTRACT

In discrete event simulation the method of control variates is often used to reduce the variance of estimation for the mean of the output response. The control variates $x_1, \ldots, x_p$ have known means and are assumed to have a joint normal distribution with the output response y. Consequently, the mean of y is estimated by linear regression.

In the present paper, it is shown that when the covariance matrix of the vector $(x_1, \ldots, x_p)$ of control variables is known, and three or more control variables are used, the usual linear regression estimator of the mean of y is one of a large class of unbiased estimators, many of which have smaller variance than the usual estimator. These estimators are shown to be adaptive to information in the data concerning the multiple correlation between the dependent variable and the control variables.

A new technical result obtained in this paper generalizes a result of Berger (1975) concerning minimax estimation of location vectors for nonnormal families of distributions.

Key words: Discrete event simulation, linear regression, adaptive estimators, minimax estimation, reduced variance, adjusted estimation of the mean, location vectors, multivariate t distribution.

[1]Research supported by NSF Grant DMS 8501966.

IMPROVED ESTIMATORS OF MEAN RESPONSE IN SIMULATION WHEN CONTROL VARIATES ARE USED

Leon Jay Gleser

Purdue University

1. Introduction

One of the main variance reduction techniques used in discrete event simulation is the method of control variates. This method attempts to exploit correlations between output responses y and certain associated auxiliary variables $x_1, x_2, \ldots, x_p$ observed during the course of each simulation run. The means $\mu_1, \mu_2, \ldots, \mu_p$ of the auxiliary variables are typically known; the goal is to estimate the mean μ_y of y.

The literature on the use of control variables in simulation is fairly recent. The first comprehensive discussion appears in Kleijen (1974). More recent surveys are Wilson (1984) and Bauer (1987).

The model underlying the use of control variables is that of linear regression with random predictors. It is assumed that n repetitions of a simulation experiment yield statistically independent observations

$$(y_i, x_{1i}, x_{2i}, \ldots, x_{pi})' \equiv (y_i, \mathbf{x}_i'), \ i = 1, 2, \ldots, n,$$

on the output response y and the vector $\mathbf{x} = (x_1, \ldots, x_p)'$ of auxiliary (control) variables. Since y and $\mathbf{x}$ result from a common set of generated random numbers and a common probabilistic structure (for example, a multiserver queue), these variables have a joint distribution with mean vector

$$\mu = (\mu_y, \mu_{\mathbf{x}}') = (\mu_y, \mu_1, \mu_2, \ldots, \mu_p)'$$

and covariance matrix

$$\Sigma = \begin{pmatrix} \sigma_{yy} & \sigma_{y\mathbf{x}} \\ \sigma_{y\mathbf{x}}' & \Sigma_{\mathbf{xx}} \end{pmatrix}.$$

The mean vector $\mu_{\mathbf{x}}$ of the control variables is known (usually from theoretical distributional information concerning these variables). It is also sometimes the case that the covariance matrix $\Sigma_{\mathbf{xx}}$ of the control variables is known.

In the literature on the use of control variables, it is usually assumed that y and $\mathbf{x}$ have a joint normal distribution. Consequently, the conditional distribution of y given $\mathbf{x} = \mathbf{x}_0$ is normal with conditional mean

$$\mu_{y|\mathbf{x}=\mathbf{x}_0} = \mu_y + \beta'(\mathbf{x}_0 - \mu_{\mathbf{x}})$$

and conditional variance

$$\sigma_{yy\cdot\mathbf{x}} = \sigma_{yy}(1 - \rho^2_{y\cdot\mathbf{x}}),$$

where

$$\beta = \sigma_{y\mathbf{x}}\Sigma^{-1}_{\mathbf{xx}}, \; \rho^2_{y\cdot\mathbf{x}} = \frac{\sigma_{y\mathbf{x}}\Sigma^{-1}_{\mathbf{xx}}\sigma'_{y\mathbf{x}}}{\sigma_{yy}},$$

are the vector of slopes for the regression of y on $\mathbf{x}$ and the squared multiple correlation (coefficient of determination), respectively.

Let

$$(\overline{y}, \overline{\mathbf{x}}') = \frac{1}{n}\sum_{i=1}^{n}(y_i, \mathbf{x}'_i)$$

and

$$W = \begin{pmatrix} w_{yy} & w_{y\mathbf{x}} \\ w_{\mathbf{x}y} & W_{\mathbf{xx}} \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^{n}(y_i - \overline{y})^2 & \sum_{i=1}^{n}(y_i - \overline{y})(\mathbf{x}_i - \overline{\mathbf{x}})' \\ \sum_{i=1}^{n}(\mathbf{x}_i - \overline{\mathbf{x}})(y_i - \overline{y}) & \sum_{i=1}^{n}(\mathbf{x}_i - \overline{\mathbf{x}})(\mathbf{x}_i - \overline{\mathbf{x}})' \end{pmatrix}$$

be the sample mean vector and sample cross-product matrix.

In the absence of data from the auxiliary variables $x_1, x_2, \ldots, x_p$, the obvious unbiased estimator of μ_y is $\overline{y}$, which has variance

$$\text{var}\,(\overline{y}) = n^{-1}\sigma_{yy}.$$

If data from the auxiliary variables is available, and the vector β of slopes is known, then

$$\overline{y}(\beta) = \overline{y} - \beta'(\overline{\mathbf{x}} - \mu_{\mathbf{x}}) \tag{1.1}$$

is an unbiased estimator of μ_y with variance

$$\text{var}\,(\overline{y}(\beta)) = n^{-1}\sigma_{yy\cdot\mathbf{x}} = n^{-1}\sigma_{yy}(1 - \rho^2_{y\cdot\mathbf{x}}).$$

The estimator $\overline{y}(\beta)$ has smaller variance than $\overline{y}$ whenever $\rho^2_{y\cdot\mathbf{x}} > 0$.

Of course, β is typically not known. In this case, we can replace β in (1.1) by the usual least squares estimator

$$\hat{\mathbf{b}} = W_{\mathbf{xx}}^{-1} w_{\mathbf{x}y} \tag{1.2}$$

of β. The resulting estimator

$$\overline{y}(\hat{\mathbf{b}}) = \overline{y} - \hat{\mathbf{b}}'(\overline{\mathbf{x}} - \mu_{\mathbf{x}}) \tag{1.3}$$

is the maximum likelihood estimator of μ_y. This estimator is unbiased for μ_y and has variance

$$\text{var}\ (\overline{y}(\hat{\mathbf{b}})) = n^{-1}(1 + \frac{p}{n-p-2})\ \sigma_{yy}(1 - \rho^2_{y\cdot\mathbf{x}}) \tag{1.4}$$

when $n \geq p+2$. (When $n < p+2$, $\overline{y}(\hat{\mathbf{b}})$ has infinite variance.) Consequently, the estimator (1.4) is superior to $\overline{y}$ (has smaller variance) as an estimator of μ_y if and only if

$$n \geq p + 2 \textit{ and } \rho^2_{y\cdot\mathbf{x}} > \frac{p}{n-2}. \tag{1.5}$$

As noted by Bauer (1987), there are usually a large number of possible control variables available in a simulation experiment. Because adding extra control variables may increase the multiple correlation $\rho_{y\cdot\mathbf{x}}$, there is the temptation to use as many of these variables as possible. Since the number n of replications of the simulation experiment is often fixed by cost considerations, this means that p may be nearly as large as n. If so, $\rho^2_{y\cdot\mathbf{x}}$ needs to be fairly large if $\overline{y}(\hat{\mathbf{b}})$ is to be superior to $\overline{y}$ as an estimate of μ_y. For example, if we can run only $n = 20$ replications and use $p = 10$ control variables, (1.5) requires $\rho^2_{y\cdot\mathbf{x}}$ to exceed 0.55.

However, $\overline{y}(\hat{\mathbf{b}})$ is not the only estimator of the form (1.1) that is unbiased for μ_y. Indeed, Theorem 1 of Section 2 shows that any function $\mathbf{b} = \mathbf{b}(W)$ of W for which $E[\mathbf{b}]$ exists yields an unbiased estimator

$$\overline{y}(\mathbf{b}) = \overline{y} - \mathbf{b}'(\overline{\mathbf{x}} - \mu_{\mathbf{x}}) \tag{1.6}$$

of μ_y. Further, it is also shown that $\overline{y}(\mathbf{b})$ has uniformly smaller variance than $\overline{y}(\hat{\mathbf{b}})$ if and only if $\mathbf{b}$ dominates $\hat{\mathbf{b}}$ in risk under the loss function

$$L(\mathbf{b};\beta,\sigma_{yy\cdot\mathbf{x}},\Sigma_{\mathbf{xx}}) = \frac{(\mathbf{b}-\beta)'\Sigma_{\mathbf{xx}}(\mathbf{b}-\beta)}{\sigma_{yy\cdot\mathbf{x}}}. \tag{1.7}$$

(The estimator $\mathbf{b}$ need not be unbiased as an estimator of β).

In this paper, we consider the special case where the covariance matrix $\Sigma_{\mathbf{xx}}$ of the vector $\mathbf{x}$ of control variables is known. Using and extending results of Berger (1975), a class of estimators of β is obtained, each of which dominates $\hat{\mathbf{b}}$ in risk when $p \geq 3, n \geq p+2$. Each member of this class thus yields an unbiased estimator $\overline{y}(\mathbf{b})$ of μ_y which has smaller variance than $\overline{y}(\hat{\mathbf{b}})$. Interestingly, estimators of the form $\overline{y}(\mathbf{b})$ can be represented as a linear combination of $\overline{y}$ and $\overline{y}(\hat{\mathbf{b}})$, with $\overline{y}$ receiving greater weight when the data indicates that $\rho^2_{y\cdot\mathbf{x}}$ is small, and $\overline{y}(\hat{\mathbf{b}})$ receiving greater weight when the data indicates that $\rho^2_{y\cdot\mathbf{x}}$ is large. That is, these estimators are adaptive.

Study of the case where $\Sigma_{\mathbf{xx}}$ is unknown is currently in progress. This problem is analytically considerably more difficult. However, Stein (1960) has shown the existence of estimators $\mathbf{b}$ of β that improve upon $\hat{\mathbf{b}}$ in risk under the loss function (1.7). The goal of our research is to provide sufficient conditions under which estimators $\mathbf{b}$ of the type considered by Stein improve upon $\hat{\mathbf{b}}$ in risk.

2. Unbiased Estimators of μ_y

The main result of this section is the following theorem.

THEOREM 1. Let $\mathbf{b} = \mathbf{b}(W)$ be any function of the sample cross-product matrix W such that $E(\mathbf{b})$ exists, all Σ. Then under the distributional assumptions for the data of Section 1,

$$\overline{y}(\mathbf{b}) = \overline{y} - \mathbf{b}'(\overline{\mathbf{x}} - \mu_{\mathbf{x}})$$

is an unbiased estimator of μ_y with variance

$$\text{var}\,(\overline{y}(\mathbf{b})) = n^{-1}\sigma_{yy}(1-\rho^2_{y\cdot\mathbf{x}})\left\{1 + \frac{E[(\mathbf{b}-\beta)'\Sigma_{\mathbf{xx}}(\mathbf{b}-\beta)]}{\sigma_{yy\cdot\mathbf{x}}}\right\}. \tag{2.1}$$

Proof. Since the data $(y_i, \mathbf{x}_i')$ are assumed to be a random sample from the $(p+1)$-variate normal distribution with mean vector μ and covariance matrix Σ, it is well known that $(\overline{y}, \overline{\mathbf{x}}')$ and W are statistically independent. Thus, since $E(\overline{y}) = \mu_y, E(\overline{\mathbf{x}} - \mu_{\mathbf{x}}) = \mathbf{0}$,

$$E[\overline{y}(\mathbf{b})] = E[\overline{y}] - E(\mathbf{b}')E(\overline{\mathbf{x}} - \mu_{\mathbf{x}}) = \mu_y,$$

for all μ_y, Σ. This shows that $\overline{y}(\mathbf{b})$ is an unbiased estimator of μ_y.

Further, it is known that

$$\overline{y} - \beta'(\overline{\mathbf{x}} - \mu_{\mathbf{x}}),\ \overline{\mathbf{x}} - \mu_{\mathbf{x}},\ W$$

are mutually statistically independent. Consequently,

$$\begin{aligned} \text{var}\,[\overline{y}(\mathbf{b})] &= \text{var}\,[\overline{y} - \beta'(\overline{\mathbf{x}} - \mu_{\mathbf{x}}) + (\mathbf{b} - \beta)'(\overline{\mathbf{x}} - \mu_{\mathbf{x}})] \\ &= \text{var}\,[\overline{y} - \beta'(\overline{\mathbf{x}} - \mu_{\mathbf{x}})] + \text{var}\,[(\mathbf{b} - \beta)'(\overline{\mathbf{x}} - \mu_{\mathbf{x}})] \\ &= n^{-1}\sigma_{yy\cdot\mathbf{x}} + E[(\mathbf{b} - \beta)'(\overline{\mathbf{x}} - \mu_{\mathbf{x}})]^2 \end{aligned} \tag{2.2}$$

since we have previously shown that the mean of $(b - \beta)'(\overline{\mathbf{x}} - \mu_{\mathbf{x}})$ is 0. However, since $\mathbf{b} = \mathbf{b}(W)$ and $\overline{\mathbf{x}}$ are independent,

$$\begin{aligned} E[(\mathbf{b} - \beta)'(\overline{\mathbf{x}} - \mu_{\mathbf{x}})]^2 &= E[(\mathbf{b} - \beta)'(\overline{\mathbf{x}} - \mu_{\mathbf{x}})(\overline{\mathbf{x}} - \mu_{\mathbf{x}})'(\mathbf{b} - \beta)] \\ &= E\left\{(\mathbf{b} - \beta)'E[\overline{\mathbf{x}} - \mu_{\mathbf{x}})(\overline{\mathbf{x}} - \mu_{\mathbf{x}})'](\mathbf{b} - \beta)\right\} \\ &= E\left[(\mathbf{b} - \beta)'\frac{1}{n}\Sigma_{\mathbf{xx}}(\mathbf{b} - \beta)\right] \\ &= \frac{\sigma_{yy\cdot x}}{n}\left\{\frac{E[(\mathbf{b} - \beta)'\Sigma_{\mathbf{xx}}(\mathbf{b} - \beta)]}{\sigma_{yy\cdot\mathbf{x}}}\right\}. \end{aligned}$$

Substituting this result into (2.2), and remembering that $\sigma_{yy\cdot\mathbf{x}} = \sigma_{yy}(1 - \rho^2_{y\cdot\mathbf{x}})$, yields (2.1). □

It follows directly from Theorem 1 that any estimator $\mathbf{b}$ of β which is a function of the data only through W, and which dominates $\hat{\mathbf{b}}$ in risk as an estimator of β under the loss function (1.7), yields an unbiased estimator $\overline{y}(\mathbf{b})$ of μ_y having smaller variance than $\overline{y}(\hat{\mathbf{b}})$.

Note. Under the distributional assumptions of Section 1, a minimal sufficient statistic for μ_y, Σ is known to be $[(\overline{y}, \overline{\mathbf{x}}'), W]$. It is an immediate consequence of Theorem 1 that this minimal sufficient statistic fails to have a complete family of distributions.

Let MVN (η, ψ) denote the multivariate normal distribution with mean vector η and covariance matrix ψ, χ^2_ν denote the chi-squared distribution with ν degrees of freedom, and $\mathcal{W}_p(\nu, \psi)$ denote the p-dimensional Wishart distribution with ν degrees of freedom and parameter ψ.

Lemma 1. Define

$$w_{yy\cdot\mathbf{x}} = w_{yy} - w_{y\mathbf{x}} W_{\mathbf{xx}}^{-1} w_{y\mathbf{x}} = w_{yy} - \hat{\mathbf{b}}' W_{\mathbf{xx}} \hat{\mathbf{b}}.$$

Then

(2.3) $$w_{yy\cdot x} \text{ is independent of } (\hat{\mathbf{b}}, W_{\mathbf{xx}})$$

(2.4) $$\hat{\mathbf{b}} | W_{\mathbf{xx}} \sim \text{ MVN } (\beta, \sigma_{yy\cdot\mathbf{x}} W_{\mathbf{xx}}^{-1})$$

(2.5) $$W_{\mathbf{xx}} \sim \mathcal{W}_p(n-1, \Sigma_{\mathbf{xx}}), \; w_{yy\cdot\mathbf{x}} \sim \sigma^2_{yy\cdot\mathbf{x}} \chi^2_{n-p-1}.$$

Proof. See Muirhead (1982, Chapter 3). □

Lemma 2. The unconditional distribution of $\hat{\mathbf{b}}$ is that of a p-variate elliptical t-distribution (Muirhead, 1982, p. 48) with $n-p$ degrees of freedom, location parameter β, and scale matrix $(n-p)^{-1}\sigma_{yy\cdot\mathbf{x}} \, \Sigma_{\mathbf{xx}}^{-1}$. That is, the density function of $\hat{\mathbf{b}}$ is

(2.6) $$f(\hat{\mathbf{b}}) = \frac{\Gamma(n/2) |\Sigma_{\mathbf{xx}}|^{\frac{1}{2}}}{\Gamma\left((n-p)/2\right) (\pi \sigma_{yy\cdot\mathbf{x}})^{p/2}} \left[1 + \frac{(\hat{\mathbf{b}} - \beta)' \Sigma_{\mathbf{xx}} (\hat{\mathbf{b}} - \beta)}{\sigma_{yy\cdot\mathbf{x}}}\right]^{-n/2}.$$

Proof. Use (2.4) and (2.5) to obtain the joint density function, $f(\hat{\mathbf{b}}, W_{\mathbf{xx}}) = f(W_{\mathbf{xx}}) \times f(\hat{\mathbf{b}} | W_{\mathbf{xx}})$, of $\hat{\mathbf{b}}$ and $W_{\mathbf{xx}}$. Let

$$\psi^{-1}(\hat{\mathbf{b}}) = \Sigma_{\mathbf{xx}}^{-1} + \sigma_{yy\cdot\mathbf{x}} (\hat{\mathbf{b}} - \beta)(\hat{\mathbf{b}} - \beta)'.$$

After some algebraic simplification, $f(\hat{\mathbf{b}}, W_{\mathbf{xx}})$ can be written as the product of $f(\hat{\mathbf{b}})$ and a function $g(W_{\mathbf{xx}}, \hat{\mathbf{b}})$ recognizable (for fixed $\hat{\mathbf{b}}$) as the density of a $\mathcal{W}_p(n, \psi(\hat{\mathbf{b}}))$ distribution. Integration over $W_{\mathbf{xx}}$, where $W_{\mathbf{xx}}$ ranges over all positive definite matrices, now completes the proof. □

Note: The proof of Lemma 2 also shows that

$$W_{\mathbf{xx}} | \hat{\mathbf{b}} \sim \mathcal{W}_p(n, \psi(\hat{\mathbf{b}})). \tag{2.7}$$

3. Estimators in the Known-$\Sigma_{\mathbf{xx}}$ Case

Suppose that the covariance matrix $\Sigma_{\mathbf{xx}}$ of the vector $\mathbf{x}$ of predictor variables is known. Consider estimators $\mathbf{b}_h$ of β of the form

$$\mathbf{b}_h = \left(1 - h\left(w_{yy \cdot \mathbf{x}}^{-1} \hat{\mathbf{b}}' \Sigma_{\mathbf{xx}} \hat{\mathbf{b}}\right)\right) \hat{\mathbf{b}}, \tag{3.1}$$

where $h(\cdot)$ is a function mapping $[0, \infty)$ to $[0, \infty)$ satisfying the following requirements:

$$\begin{aligned} &\text{(i) } h(u) \text{ is nonincreasing in } u \geq 0, \\ &\text{(ii) } r(u) \equiv uh(u) \text{ is nondecreasing in } u \geq 0, \end{aligned} \tag{3.2}$$

THEOREM 2. Assume that $p \geq 3, n \geq p + 2$. Let the estimator $\mathbf{b}_h$ of β be defined by (3.1) and (3.2). If

$$r(u) = uh(u) \leq \frac{2(p-2)}{(n-p)(n-p-1)}, \tag{3.3}$$

then $\mathbf{b}_h$ has risk everywhere (over $\beta, \sigma_{yy \cdot \mathbf{x}}$) at least as small as that of $\hat{\mathbf{b}}$ under the loss function (1.7). Consequently, $\overline{y}(\mathbf{b}_h)$ is an unbiased estimator of μ_y having variance everywhere less than or equal to the variance of $\overline{y}(\hat{\mathbf{b}})$.

Berger (1975, Example 3) notes that the p-variate elliptical t-densities (which he calls "Cauchy-like") can be represented as scale mixtures of p-variate normal densities. If $\sigma_{yy \cdot \mathbf{x}}$ were known, and $w_{yy \cdot \mathbf{x}}$ in (3.1) replaced by $\sigma_{yy \cdot \mathbf{x}}$, Theorem 2 would be a direct application of Theorem 1 of Berger. Since $\sigma_{yy\cdot\mathbf{x}}$ is not known and must be estimated, proof of Theorem 2 requires a slight generalization of Berger's result.

Generalization of Theorem 1 of Berger (1975)

Let the p-dimensional random vector $\mathbf{z}$ have density

$$f(\mathbf{z}) = \int_0^\infty \frac{|\Delta|^{-\frac{1}{2}}}{(2\pi\tau^2 v)^{p/2}} \cdot \exp\left[-\frac{(\mathbf{z}-\theta)'\Delta^{-1}(\mathbf{z}-\theta)}{2\tau^2 v}\right] dF(v) \tag{3.4}$$

where $\theta : p \times 1$ and $\tau^2 > 0$ are unknown parameters, where Δ is a known positive definite matrix, and where $F(\cdot)$ is any known cdf on $(0, \infty)$. Let w be a scalar random variable distributed, independently of $\mathbf{z}$, with the property that $\tau^{-2}w$ has a known distribution on $(0, \infty)$. [Thus, τ^2 is a scale-parameter for the distribution of w.] Consider estimating θ under the quadratic loss function

$$L(\delta; \theta, \tau^2) = \frac{(\delta-\theta)'Q(\delta-\theta)}{\tau^2}, \tag{3.5}$$

where Q is a known positive definite matrix. Since Berger (1975) shows that $\mathbf{z}$ is a minimax estimator of θ in the known-τ^2 case, $\mathbf{z}$ is also minimax in the present problem. To improve upon $\mathbf{z}$ in risk then requires $p \geq 3$, so we henceforth assume that this is the case.

In the known-τ^2 case, Berger's Σ corresponds to our $\tau^2\Delta$, and Berger's Q to our $\tau^{-2}Q$. Thus, the class of estimators shown by Berger to dominate $\mathbf{z}$ in the known-τ^2 case has the form

$$\left(I_p - h\left(\frac{z'\Delta^{-1}Q^{-1}\Delta^{-1}z}{\tau^2}\right) Q^{-1}\Delta^{-1}\right)\mathbf{z}.$$

Since τ^2 is unknown, we replace τ^2 by w. (We could instead replace τ^2 by cw for any positive constant c, for example $c = 1/E_{\tau^2=1}(w)$, but in that case we merely redefine $h(\cdot)$ to absorb the constant c.) Thus, we consider estimators of the form

$$\delta_h(\mathbf{z}, w) = \left[I_p - h\left(\frac{\mathbf{z}'\Delta^{-1}Q^{-1}\Delta^{-1}\mathbf{z}}{w}\right) Q^{-1}\Delta^{-1}\right]\mathbf{z}, \tag{3.6}$$

where the function $h(\cdot)$ maps $[0, \infty)$ to $[0, \infty)$ and satisfies (3.2).

THEOREM 3. Suppose that when $\theta = 0$ and $\tau^2 = 1$, $E(\mathbf{z}'\mathbf{z}), E(\mathbf{z}'\mathbf{z})^{-1}, E(w^{-1})$ are all finite. Let

$$d = E_{\theta=0,\tau^2=1}\left[(\mathbf{z}'\Delta^{-1}\mathbf{z})^{-1}\right] E_{\tau^2=1}(w).$$

Then, if

$$(3.7) \qquad r(u) = uh(u) \le 2(d)^{-1}$$

for all $u \ge 0$, the estimator $\delta_h(\mathbf{z}, w)$ has risk everywhere (over θ, τ^2) less than or equal to the risk of $\mathbf{z}$, and hence is minimax.

Proof. Note that

$$\delta_h(\tau^{-1}\mathbf{z}, \tau^{-2}w) = \tau\delta_h(\mathbf{z}, w).$$

so that for all h (including $h(\cdot) \equiv 0$, yielding $\delta_0(\mathbf{z}, w) = \mathbf{z}$),

$$L(\delta_h(\mathbf{z}, w); \theta, \tau^2, \Delta) = L(\delta_h(\tau^{-1}\mathbf{z}, \tau^{-2}w); \tau^{-1}\theta, 1, \Delta).$$

That is, the estimation problem for the class of estimators $\delta_h(\mathbf{z}, w)$ is invariant under the transformation $\mathbf{z} \to \tau^{-1}\mathbf{z}$, $\theta \to \tau^{-1}\theta$. Consequently, we can assume without loss of generality that $\tau^2 = 1$.

As in Berger (1975), let $A_1 \ge A_2 \ge \ldots \ge A_p \ge 0$ be the (ordered) latent roots of ΔQ^{-1}. There exists a nonsingular matrix B such that $B'QB = I_p$, $B'\Sigma^{-1}B = D_A^{-1}$, where

$$D_A = \text{diag}(A_1, \ldots, A_p).$$

Transforming

$$\mathbf{z} \to B^{-1}\mathbf{z}, \ \theta \to B^{-1}\theta,$$

produces a "canonical" estimation problem in which the distribution of $\mathbf{z}$ has parameter $\Delta = D_A$, the loss function (3.5) has centering matrix $Q = I_p$, and the estimators $\delta_h(\mathbf{z}, w)$ have the form

$$\delta_h(\mathbf{z}, w) = \left[I_p - h\left(\frac{\mathbf{z}'D_A^{-2}\mathbf{z}}{w}\right) D_A^{-1}\right] \mathbf{z}.$$

Let

$$(3.8) \qquad \omega(\theta) = E_\theta\left[(\mathbf{z} - \theta)'(\mathbf{z} - \theta) - (\delta_h(\mathbf{z}, w) - \theta)'(\delta_h(\mathbf{z}, w) - \theta)\right]$$

be the difference in risks between $\mathbf{z}$ and $\delta_h(\mathbf{z}, w)$. (Remember that τ^2 is assumed to equal 1). We need to show that $\omega(\theta) \ge 0$ for all θ.

Since w and $\mathbf{z}$ are independent, we can fix w and take expectation in (3.8) in the order $E_w E_{\mathbf{z}}$. Let $h_w(u) = h(uw^{-1})$. Using (3.2), it is straightforward to show that $h_w(u)$ is nonincreasing in $u, u \geq 0$, and that $r_w(u) = uh_w(u)$ is nondecreasing in $u, u \geq 0$. Further,

$$\delta_h(\mathbf{z}, w) = (I_p - h_w(\mathbf{z}'D_A^{-2}\mathbf{z})D_A^{-1})\mathbf{z}.$$

Finally, if $c = \sup_{u\geq 0} r(u)$, then

$$\sup_{u\geq 0} r_w(u) = cw.$$

Consequently, we can follow the steps on pp. 1320–1322 in Berger (1975) to obtain

$$\omega(\theta) \geq E_w\left[\left(\int_0^\infty [2(p-2) - \frac{cw}{v}]dF(v)\right) T(w,\theta)\right],$$

where

$$T(w,\theta) = \int_0^\infty \int_{R^p} \frac{v^{-\frac{p-2}{2}} h(\frac{\mathbf{z}'D_A^{-2}\mathbf{z}}{w}) \exp\{-\frac{1}{2v}(\mathbf{z}-\theta)'D_A^{-1}(\mathbf{z}-\theta)\}d\mathbf{z}dF(v)}{(2\pi)^{p/2}\prod_{i=1}^{p} A_i^{1/2}}.$$

Since $h(u)$ is nonincreasing in $u \geq 0$, $T(w,\theta)$ is nondecreasing in w for fixed θ. Consequently,

$$\omega(\theta) \geq E_w\left[\left(\int_0^\infty [2(p-2) - \frac{cw}{v}]dF(v)\right)\right] E_w[T(w,\theta)]. \tag{3.9}$$

Since $T(w,\theta) \geq 0$, all $w \geq 0$, all θ, in order for $w(\theta)$ to be greater than or equal to 0 for all θ it is sufficient that

$$\begin{aligned} 0 &\leq E_w\left[\int_0^\infty [2(p-2) - \frac{cw}{v}]dF(v)\right] \\ &= 2(p-2) - cE(w)E(v^{-1}) \end{aligned}$$

or equivalently

$$c \leq \frac{2(p-2)}{E(w)E(v^{-1})}. \tag{3.10}$$

When $\theta = 0, v\mathbf{z}'D_A\mathbf{z} \sim \chi_p^2$ and is independent of v. Thus

$$E_{\theta=0}\left[(\mathbf{z}'D_A^{-1}\mathbf{z})^{-1}\right] = E\left[\frac{1}{v}\right] E\left[\frac{1}{\chi_p^2}\right] = \frac{E(v^{-1})}{p-2},$$

and we conclude that

$$c = \sup_{u \geq 0} r(u) \leq \frac{2}{E(w) E_{\theta=0}(\mathbf{z}' D_A^{-1} \mathbf{z})} \tag{3.11}$$

is sufficient for $\omega(\theta) \geq 0$, and hence for $\delta_h(\mathbf{z}, w)$ to dominate $\mathbf{z}$ in risk. Transforming back from the canonical form of the estimation problem to our original formulation yields (3.7) as a sufficient condition for $\delta_h(\mathbf{z}, w)$ to have risk everywhere less than or equal to $\mathbf{z}$. □

Proof of Theorem 2

Make the following correspondences between the notation of Theorem 2 and that of Theorem 3:

$$\hat{\mathbf{b}} \leftrightarrow \mathbf{z}, \ \Sigma_{\mathbf{xx}}^{-1} \leftrightarrow \Delta, \ \beta \leftrightarrow \theta, \ \sigma_{yy\cdot\mathbf{x}} \leftrightarrow \tau^2, \ w_{yy\cdot\mathbf{x}} \leftrightarrow w,$$

Further, the loss function used in Theorem 2 has $Q = \Sigma_{\mathbf{xx}}$.

The density (2.6) of $\hat{b}$ can be written in the form (3.3) with $F(v)$ being the cdf of $v \sim (\chi^2_{n-p})^{-1}$. Note that from (2.5), $w_{yy\cdot\mathbf{x}} \sim \sigma_{yy\cdot\mathbf{x}}\, \chi^2_{n-p-1}$, so that $E[w_{yy\cdot\mathbf{x}}] = (n-p-1)$ when $\sigma_{yy\cdot\mathbf{x}} = 1$. Further, if $v^{-1} \sim \chi^2_{n-p}$, then $E(v^{-1}) = n-p$. Making use of the condition (3.10), which is equivalent to (3.7), yields the condition (3.3).

Adaptive Estimation

It was noted in Section 1 that $\overline{y}(\hat{\mathbf{b}})$ has smaller variance than $\overline{y}$ as an estimator of μ_y only when $\rho^2_{y\cdot\mathbf{x}}$ is large. In particular, we would want to use $\overline{y}$ in preference to $\overline{y}(\hat{\mathbf{b}})$ as an estimator of μ_y when $\rho^2_{y\cdot\mathbf{x}} = 0$ (equivalently, $\beta = 0$). An appropriate (likelihood ratio) test statistic for testing $H_0 : \rho^2_{y\cdot\mathbf{x}} = 0$ when $\Sigma_{\mathbf{xx}}$ is known is

$$T = \frac{\hat{\mathbf{b}}' \Sigma_{\mathbf{xx}} \hat{\mathbf{b}}}{w_{yy\cdot\mathbf{x}}}.$$

We reject H_0 for sufficiently large values of T. Thus, we might naively try to use the value of T to choose between $\overline{y}$ and $\overline{y}(\hat{\mathbf{b}})$ as estimators of μ_y. Let

$$\mathbf{b}_h = \left(1 - h\left(\frac{\hat{\mathbf{b}}' \Sigma_{\mathbf{xx}} \hat{\mathbf{b}}}{w_{yy\cdot\mathbf{x}}}\right)\right) \hat{\mathbf{b}} = (1 - h(T))\, \hat{\mathbf{b}}$$

be any of the estimators (3.1). Then

$$\begin{aligned}
\overline{y}(\mathbf{b}_h) &= \overline{y} - \mathbf{b}_h'(\overline{\mathbf{x}} - \mu_{\mathbf{x}}) \\
&= [h(T) + (1 - h(T))]\, \overline{y} - (1 - h(T))\hat{\mathbf{b}}'(\overline{\mathbf{x}} - \mu_{\mathbf{x}}) \qquad (3.12) \\
&= h(T)\overline{y} + (1 - h(T))\, \overline{y}(\hat{\mathbf{b}}).
\end{aligned}$$

Since $h(T)$ is nonincreasing in T, large values of T (indicating that $H_0 : \rho^2_{y\cdot\mathbf{x}} = 0$ may be false) cause greater weight to be placed on $\overline{y}(\hat{\mathbf{b}})$ in (3.12). Small values of T cause greater weight to be placed on $\overline{y}$. Indeed, T is stochastically increasing in $\rho^2_{y\cdot\mathbf{x}}$. Consequently, each member of the class of estimators $\overline{y}(\mathbf{b}_h)$ defined by (3.1) and (3.2) is adaptive to the information provided by the data concerning the magnitude of $\rho^2_{y\cdot\mathbf{x}}$. This property of these estimators in part explains why some of these estimators have lower variance than $\overline{y}(\hat{\mathbf{b}})$.

To obtain some idea of the magnitude of improvement that can be achieved, consider the somewhat crude estimator

$$\mathbf{b}^* = \left(1 - \frac{p-2}{(n-p)(n-p-1)T}\right)\hat{\mathbf{b}}, \qquad (3.13)$$

which nevertheless dominates $\hat{\mathbf{b}}$ in risk. It can be shown directly that the difference in risks between $\hat{\mathbf{b}}$ and $\mathbf{b}^*$ as estimators of β when $\beta = 0$ is given by

$$(p-2)E_{\beta=0,\sigma_{yy\cdot\mathbf{x}}=1}\left[\frac{p-2}{(n-p)(n-p-1)T}\right] = \frac{(p-2)}{n-p}.$$

Consequently, the risk of $\mathbf{b}^*$ when $\beta = 0$ is

$$\frac{p}{n-p-2} - \frac{p-2}{n-p} = \frac{2(n-2)}{(n-p)(n-p-2)}$$

and the variance of $\overline{y}(\mathbf{b}^*)$ when $\beta = 0$ is equal to

$$n^{-1}\sigma_{yy}\left(1 + \frac{2(n-2)}{(n-p)(n-p-2)}\right).$$

When the number of replications n is even modestly large, and $\beta = 0$, the improvement in variance of $\overline{y}(\mathbf{b}^*)$ over $\overline{y}(\hat{\mathbf{b}})$ is small, mainly because of the n^{-1} term appearing in both variances. The *relative improvement* in variance, however, is impressive. Indeed, when $\beta = 0$,

$$\frac{\text{var}\left[\overline{y}(\hat{\mathbf{b}})\right] - \text{var}\left[\overline{y}(\mathbf{b}^*)\right]}{\text{var}\left[\overline{y}(\hat{\mathbf{b}})\right]} = \frac{(n-p-1)(p-2)}{(n-1)(n-p)}. \tag{3.14}$$

For example, when $n = 20, p = 10$, the right-hand side of (3.14) is 0.38, indicating a 38% reduction in variance. When $n = 40, p = 25$, there is a 55% reduction in variance for $\overline{y}(\mathbf{b}^*)$ versus $\overline{y}(\hat{\mathbf{b}})$.

REFERENCES

Bauer, Kenneth W., Jr. (1987). *Control Variate Selection for Multiresponse Simulation.* Ph.D. Dissertation, Department of Industrial Engineering, Purdue University.

Berger, James (1975). Minimax estimation of location vectors for a wide class of densities. *Ann. Statist.* **3**, 1318–1328.

Kleijen, J. P. C. (1975). *Statistical Techniques in Simulation, Part 1 and 2.* Marcel Dekker:New York.

Muirhead, Robb J. (1982). *Aspects of Multivariate Statistical Theory.* John Wiley:New York.

Stein, Charles (1960). Multiple regression. In *Contributions to Probability and Statistics. Essays in Honor of Harold Hotelling.* (Olkin, I.; Ghurye, S. G.; Hoeffding, W.; Madow, W. G.; Mann, H. B., eds). Stanford University Press:Stanford, California.

Wilson, James R. (1984). Variance reduction techniques for digital simulation. *Amer. J. of Math. and Man. Sci.* **1**, 227–312.