(2)

NPS55-90-09

NAVAL POSTGRADUATE SCHOOL

Monterey, California

VARIANCE REDUCTION FOR QUANTILE ESTIMATES IN SIMULATIONS VIA NONLINEAR CONTROLS

Peter A. W. Lewis
and
Richard L. Ressler

April 1990

Prepared for:
Chief of Naval Research
Arlington, VA

NAVAL POSTGRADUATE SCHOOL
MONTEREY, CALIFORNIA

Rear Admiral R. W. West, Jr.
Superintendent

Harrison Shull
Provost

This report was sponsored by the Chief of Naval Research and funded by the Naval Postgraduate School.

This report was prepared by:

P. A. W. LEWIS
Professor of Operations Research

Reviewed by:

PETER PURDUE
Professor and Chairman
Department of Operations Research

Released by:

Dean of Faculty and Graduate Studies

UNCLASSIFIED

Security Classification of this page

REPORT DOCUMENTATION PAGE

1a Report Security Classification: Unclassified	1b Restrictive Markings
2a Security Classification Authority	3 Distribution Availability of Report: Approved for public release; distribution is unlimited.
2b Declassification/Downgrading Schedule	
4 Performing Organization Report Number(s): NPS55-90-09	5 Monitoring Organization Report Number(s)
6a Name of Performing Organization: Naval Postgraduate School; 6b Office Symbol (*If Applicable*): OR	7a Name of Monitoring Organization: Chief of Naval Research
6c Address (*city, state, and ZIP code*): Monterey, CA 93943-5000	7b Address (*city, state, and ZIP code*): Arlington, VA
8a Name of Funding/Sponsoring Organization: Naval Postgraduate School; 8b Office Symbol (*If Applicable*)	9 Procurement Instrument Identification Number: O&MN, Direct Funding
8c Address (*city, state, and ZIP code*)	10 Source of Funding Numbers: Program Element Number; Project No; Task No; Work Unit Accession No

11 Title (*Include Security Classification*): Variance Reduction for Quantile Estimates in Simulations via Nonlinear Controls

12 Personal Author(s): Peter A. W. Lewis and Richard L. Ressler

13a Type of Report	13b Time Covered	14 Date of Report (*year, month,day*)	15 Page Count
Technical	From To	1990, April	33

16 Supplementary Notation: The views expressed in this paper are those of the author and do not reflect the official policy or position of the Department of Defense or the U.S. Government.

17 Cosati Codes		
Field	Group	Subgroup

18 Subject Terms (*continue on reverse if necessary and identify by block number*): Variance reduction; quantiles; nonlinear controls; transformations; ACE; least-squares regression; jackknifing. (ICR)

19 Abstract (*continue on reverse if necessary and identify by block number*)

Linear controls are a well known simple technique for achieving variance reduction in computer simulation. Unfortunately the effectiveness of a linear control depends upon the correlation between the statistic of interest and the control, which is often low. Since statistics often have a nonlinear relationship with the potential control variables, nonlinear controls offer a means for improvement over linear controls. This paper focuses on the use of nonlinear controls for reducing the variance of quantile estimates in simulation. It is shown that one can substantially reduce the analytic effort required to develop a nonlinear control from a quantile estimator by using a strictly monotone transformation to create the nonlinear control. It is also shown that as one increases the sample size for the quantile estimator, the asymptotic multivariate normal distribution of the quantile of interest and the control reduces the effectiveness of the nonlinear control to that of the the linear control. However, the data has to be sectioned to obtain an estimate of the variance of the controlled quantile estimate. Graphical methods are suggested for selecting the section size that maximizes the effectiveness of the nonlinear control.

Keywords:

20 Distribution/Availability of Abstract	21 Abstract Security Classification
[X] unclassified/unlimited [ ] same as report [ ] DTIC users	Unclassified

22a Name of Responsible Individual	22b Telephone (*Include Area code*)	22c Office Symbol
Lewis, P. A. W.	(408) 646-2283	OR/Lw

DD FORM 1473, 84 MAR — 83 APR edition may be used until exhausted — All other editions are obsolete

security classification of this page: UNCLASSIFIED

VARIANCE REDUCTION FOR QUANTILE ESTIMATES IN SIMULATIONS VIA NONLINEAR CONTROLS

Richard L. Ressler Peter A. W. Lewis

Naval Postgraduate School
Monterey, CA 93943

Key Words and Phrases: variance reduction; quantiles; nonlinear controls; transformations; ACE; least-squares regression; jackknifing

ABSTRACT

Linear controls are a well known simple technique for achieving variance reduction in computer simulation. Unfortunately the effectiveness of a linear control depends upon the correlation between the statistic of interest and the control, which is often low. Since statistics often have a nonlinear relationship with the potential control variables, nonlinear controls offer a means for improvement over linear controls. This paper focuses on the use of nonlinear controls for reducing the variance of quantile estimates in simulation. It is shown that one can substantially reduce the analytic effort required to develop a nonlinear control from a quantile estimator by using a strictly monotone transformation to create the nonlinear control. It is also shown that as one increases the sample size for the quantile estimator, the asymptotic multivariate normal distribution of the quantile of interest and the control reduces the effectiveness of the nonlinear control to that of the linear control. However, the data has to be sectioned to obtain an estimate of the variance of the controlled quantile estimate. Graphical methods are suggested for selecting the section size that maximizes the effectiveness of the nonlinear control.

1 OUTLINE OF THE PAPER

The paper begins with a short discussion of quantiles and the properties of a quantile estimator, with emphasis on the need for a reliable estimator for the variance of the quantile estimator. The next part of the paper discusses linear controls for quantile estimates and the subtleties involved with estimating the coefficients for the control functions. The discussion of linear controls is followed by a discussion of nonlinear controls and their application to reducing the variance of quantile

estimates for a fixed simulation sample size. The final part of the paper presents an extract of results from a simulation experiment where crude, linearly controlled and nonlinearly controlled estimators are compared. Throughout the paper the emphasis is on quantile estimation for continuous random variables, though other cases are of interest.

2 QUANTILES

2.1 Properties of a Quantile Estimator

Let Y be a random variable with a right-continuous distribution function defined by

$$F_Y(y) = \Pr\{Y \leq y\}, \quad -\infty < y < \infty.$$

Following Serfling (1980) define the α quantile of Y, y_α, for $0 < \alpha < 1$, as the value

$$F_Y^{-1}(\alpha) = \inf\{y : F_Y(y) \geq \alpha\}. \tag{1}$$

If $F_Y(y)$ is strictly increasing, y_α is unique for each α. Additional restrictions on $F_Y(y)$, such as continuity at y_α, may be needed for the existence of certain asymptotic properties and will be stated as required.

Given a simulation sample of n independent and identically distributed (i.i.d.) samples of Y, namely $Y_1, \ldots, Y_n$, one can construct a sample distribution function, F_n, by placing at each observation Y_i, a mass $1/n$. Thus F_n may be represented as

$$F_n(y) = \frac{1}{n}\sum_{i=1}^{n} \mathrm{I}(Y_i \leq y), \quad -\infty < y < \infty$$

where $\mathrm{I}(\cdot)$ is an indicator function which returns 1 if the argument is true and 0 otherwise.

For a sample of size n, one can define a nonparametric estimator of the α quantile, $\hat{y}_\alpha(n)$, as the sample α quantile of the sample distribution function, or

$$\hat{y}_\alpha(n) = F_n^{-1}(\alpha).$$

Using the sample α quantile to estimate y_α is equivalent to using the order statistics of the sample, $Y_{(1)} <, \ldots, < Y_{(n)}$, and defining a nonparametric estimator of the α quantile, $\hat{y}_\alpha(n)$, as in Lewis and Orav (1989), as

$$\hat{y}_\alpha(n) = Y_{(r)} = \begin{cases} Y_{(n\alpha)} & \text{if } n\alpha \text{ is an integer} \\ Y_{(\lfloor n\alpha \rfloor + 1)} & \text{if } n\alpha \text{ is not an integer} \end{cases} \tag{2}$$

where $\lfloor w \rfloor$ denotes the integral part of w.

For a given n and α, $\hat{y}_\alpha(n)$ is the rth order statistic from the n-sized sample where r is determined as in (2). The following results on the distribution of $\hat{y}_\alpha(n)$ are well known (David 1970, chap. 1-3 or Kendall and Stuart 1977, pp. 251-252).

Let $F_{\hat{y}_\alpha(n)}(y)$ be the cumulative distribution function of the quantile estimator. Then $F_{\hat{y}_\alpha(n)}(y)$ can be written as

$$\begin{aligned} F_{\hat{y}_\alpha(n)}(y) &= \Pr\{\hat{y}_\alpha(n) \le y\} \\ &= \Pr\{\text{ at least } r \text{ of the } n\ Y_i \text{ are} \le y\} \\ &= \sum_{i=r}^{n} \binom{n}{i} F_Y^i(y)[1 - F_Y(y)]^{n-i}, \end{aligned} \tag{3}$$

since the term in the summand is the binomial probability that exactly i of the Y_i are less than or equal to y. If the Y_i are continuous with a density function $f_Y(y)$, the density function of $\hat{y}_\alpha(n)$ is

$$f_{\hat{y}_\alpha(n)}(y) = \frac{1}{\mathrm{B}_{(r,n-r+1)}} F_Y^{r-1}(y)[1 - F_Y(y)]^{n-r} f_Y(y)$$

where $\mathrm{B}(\cdot,\cdot)$ represents the complete beta function. Unfortunately, while $\hat{y}_\alpha(n)$ is a nonparametric estimator, (3) shows that the distribution of the quantile estimator $\hat{y}_\alpha(n)$ depends not only on n and α but also on the unknown distribution of the underlying Y.

The bias and variance of $\hat{y}_\alpha(n)$ also depend on n, α, and the distribution of the underlying Y. Assume that $F_Y(y)$ is continuous with a density function $f_Y(y)$ which is differentiable and nonzero at y_α. The following result for the expected value of the quantile estimator can be derived from results in David (1970, p. 65):

$$E[\hat{y}_\alpha(n)] = y_\alpha - \frac{\epsilon}{n f_Y(y_\alpha)} - \frac{\alpha(1-\alpha)}{2(n+2)} \frac{f_Y'(y_\alpha)}{f_Y^2(y_\alpha)} + \mathrm{O}\left(\frac{1}{n^2}\right), \tag{4}$$

where ϵ is a sawtooth function of n and α such that $|\epsilon| < 1$ and $f'(\cdot)$ denotes the derivative of the function $f(\cdot)$. An expansion for the variance of the quantile estimator can be derived in similar fashion as

$$\mathrm{var}[\hat{y}_\alpha(n)] = \sigma^2_{\hat{y}_\alpha(n)} = \frac{\alpha(1-\alpha)}{(n+2) f_Y^2(y_\alpha)} + O\left(\frac{1}{n^2}\right). \tag{5}$$

The notation $g(n) = \mathrm{O}(1/n^2)$ means that the absolute value of $g(n)/(1/n^2)$ remains bounded as n goes to infinity.

There are also well known asymptotic results for $\hat{y}_\alpha(n)$ (Serfling, 1980, sec. 2.3).

- If y_α is the unique solution y of $F(y-) \le \alpha \le F(y)$, then $\hat{y}_\alpha(n) \longrightarrow y_\alpha$ with probability 1 as $n \longrightarrow \infty$.
- If $F_Y(y)$ possesses a density $f_Y(y)$ in a neighborhood of y_α, and $f_Y(y)$ is positive and continuous at y_α, then $\hat{y}_\alpha(n)$ has an asymptotic normal distribution in that

$$F_{\hat{y}_\alpha(n)}(y) \sim \mathrm{N}\left\{ y_\alpha, \left(\frac{\alpha(1-\alpha)}{n f_Y^2(y_\alpha)}\right)^{1/2} \right\} \quad \text{as } n \longrightarrow \infty.$$

- Weiss (1964) proved that under mild conditions, the sample marginal quantiles from a multivariate population with an absolutely continuous joint distribution function have an asymptotic multivariate normal distribution. The asymptotic covariance is a function of the multivariate distribution of the underlying multivariate population. This multivariate result is important because of the role of the joint distribution of the controlled and controlling statistics in the theory of controls for variance reduction.

2.2 Using Sectioning to Estimate the Variance of a Quantile Estimator

When using (2) to calculate a point estimate of the α quantile, one must also estimate the variance or equivalently the standard deviation of the point estimate. One could estimate the density of Y at y_α and use (5) to estimate the variance. However, the instability of density estimates at extreme quantiles can cause this to be a very biased and unstable estimate of the variance of $\hat{y}_\alpha(n)$. A more general technique is to use sectioning to calculate both a point estimate of the quantile and an estimate of the variance of the point estimate. While non-parametric confidence intervals are available for crude quantile estimates (see Mood Graybill and Boes 1974, p. 312), the confidence intervals are not appropriate for controlled estimates. A brief discussion of sectioning follows; for a detailed discussion of sectioning see Lewis and Orav (1989, chap. 9).

Let the random variable $\hat{y}_\alpha(n)$ be the function of independent and identically distributed random variables $Y_1, \ldots, Y_n$ defined in (2) such that $\hat{y}_\alpha(n)$ is a point estimator of y_α. Let $\sigma^2_{\hat{y}_\alpha(n)}$ denote the variance of $\hat{y}_\alpha(n)$. Assume for now that there are a total of $N = m \times n$ independent samples of Y, namely $Y_1, \ldots, Y_n, \ldots, Y_N$. The sectioned point estimator, $\overline{\hat{y}_\alpha}(m,n)$, is constructed as follows:

1. Divide the N samples of the random variable Y into m sections with n samples each where for simplicity $n \times m = N$ (equivalently, replicate a sample of size n, m times).

2. For the jth section, $j = 1, \ldots, m$, use (2) to compute $\hat{y}_{\alpha,j}(n)$.

3. Compute $\overline{\hat{y}_\alpha}(m,n)$ as:

$$\overline{\hat{y}_\alpha}(m,n) = \frac{1}{m} \sum_{j=1}^{m} \hat{y}_{\alpha,j}(n). \tag{6}$$

   The point estimator $\overline{\hat{y}_\alpha}(m,n)$ is a sample mean of m independent estimates, each of which is based on n samples.

4. Estimate the variance of $\overline{\hat{y}_\alpha}(m,n)$, namely $\sigma^2_{\overline{\hat{y}_\alpha}(m,n)}$, with the sample variance of the sample mean:

$$S^2_{\overline{\hat{y}_\alpha}(m,n)} = \frac{1}{m(m-1)} \sum_{j=1}^{m} \left\{ \hat{y}_{\alpha,j}(n) - \overline{\hat{y}_\alpha}(m,n) \right\}^2. \tag{7}$$

One advantage of sectioning to estimate the variance of the quantile estimate over estimating the density is that since the $\hat{y}_{\alpha,j}(n)$ in step 2 above are i.i.d. and the point estimator $\overline{\hat{y}_\alpha}(m,n)$ is their sample mean, $S^2_{\overline{\hat{y}_\alpha}(m,n)}$ is an unbiased estimate of the variance of the point estimate. Furthermore, if the $\hat{y}_{\alpha,j}(n)$ are approximately normally distributed, one can develop approximate confidence intervals for $\overline{\hat{y}_\alpha}(m,n)$ based on a t-statistic with $m-1$ degrees of freedom. A disadvantage of sectioning is the increase in the bias of the point estimate; the first-order bias predicted by (4) for $\overline{\hat{y}_\alpha}(m,n)$ is m times that for $\hat{y}_\alpha(N)$, a point estimate based on all N samples.

For fixed N, the selection of m and n involves a tradeoff between the bias and the variance of $\overline{\hat{y}_\alpha}(m,n)$as well as the precision of the estimate of the variance of $\overline{\hat{y}_\alpha}(m,n)$. To minimize the bias in $\overline{\hat{y}_\alpha}(m,n)$, as well as improve the approximation to normality of the individual $\hat{y}_j(n)$, one would like n to be large. A drawback of increasing n is the decrease in precision of the estimate of the variance of the point estimate as well as a decrease in the degrees of freedom, $m-1$, for the t-statistic, which relaxes the confidence interval. Using (5) and (7), one can write the expansion for the variance of the sectioned estimate in terms of m only as

$$\sigma^2_{\overline{\hat{y}_\alpha}(m,n)} = \frac{\sigma^2_{\hat{y}_\alpha(n)}}{m} = \frac{\beta}{(N+2m)} + \frac{m\gamma}{N^2} + O\left(\frac{1}{N^2}\right), \tag{8}$$

where β and γ are constants determined by $F_Y(y)$ and α. The presence of m in both the denominator and the numerator in (8) implies, for fixed N, that the value of m which minimizes the variance is a function of the relative magnitudes of β and γ. If β is small relative to γ, one should choose a small m in order to minimize the variance. The value for m must be at least 2 in order to use (7) to estimate the variance. Values for m and n which will minimize the variance or the mean square error of the point estimate can be determined as functions of terms such as β and γ. However, these terms are in turn functions of the distribution of Y which is unknown. After consideration of the above, Lewis and Orav (1989, p. 262) suggest as a "rough rule of thumb" to make m between 12 and 20 for samples with N over 1000. This usually gives sufficient precision for the estimate of the variance of $\overline{\hat{y}_\alpha}(m,n)$.

Once m and n have been selected, the variance of the point estimate can be estimated. Equation (5) shows that $\sigma^2_{\hat{y}_\alpha(n)}$ is a decreasing function of n. For fixed m, a decrease in $\sigma^2_{\hat{y}_\alpha(n)}$ will cause a corresponding decrease in $\sigma^2_{\overline{\hat{y}_\alpha}(m,n)}$. A technique for reducing $\sigma^2_{\hat{y}_\alpha(n)}$ without increasing n is linear controls.

3 LINEAR CONTROL OF QUANTILES

3.1 Single and Multiple Linear Controls

3.1.1 A Single Linear Control

Linear controls is a variance reduction technique which can be used to reduce the variance of an estimate of a statistic of interest, often a sample mean. The statistic of interest in this paper is the quantile estimator $\hat{y}_\alpha(n)$ from (2) and eventually the individual section estimate $\hat{y}_{\alpha,j}(n)$ from (6).

To use a linear control for variance reduction a random variable generated in the simulation, called the control or control variable, which is correlated with $\hat{y}_\alpha(n)$, must be available. The expected value of the control must be known, either exactly or approximately. Let C be a random variable which is generated via simulation. Although an estimator of the α quantile of C is not necessarily the most effective control for a given quantile or Y, for purposes of discussion we will use as the control variable the estimator of the α quantile of C as defined in (2), namely $\hat{c}_\alpha(n)$. The random variable $\hat{c}_\alpha(n)$ is a function of n i.i.d. samples of the random variable C. If $\hat{c}_\alpha(n)$ is generated as part of the simulation that produces the samples of Y it will be called an internal control variable. If $\hat{c}_\alpha(n)$ is generated as output from a different simulation, it will be called an external control variable.

The linear control scheme for variance reduction, with a single control, uses as a control function a linear additive combination of the control and its expected value to produce a controlled estimate $\hat{y}'_\alpha(n)$ where the prime applied to an estimate implies that it is a controlled estimate. The control function, with coefficient θ, is subtracted off from the uncontrolled or crude estimate $\hat{y}_\alpha(n)$ to produce the controlled estimate as follows:

$$\hat{y}'_\alpha(n) = \hat{y}_\alpha(n) - \theta \left\{ \hat{c}_\alpha(n) - \mathrm{E}[\hat{c}_\alpha(n)] \right\}. \tag{9}$$

Putting aside the question of sectioning for now, the purpose of using a control is to minimize the variance of the controlled estimate, $\sigma^2_{\hat{y}'_\alpha(n)}$, for a fixed sample size n. If the statistic of interest is $\hat{y}_{\alpha,j}(n)$ from (6), minimizing its variance will, for fixed m, minimize the variance of the section estimate $\overline{\hat{y}_\alpha}(m,n)$. The value of θ which minimizes $\sigma^2_{\hat{y}'_\alpha(n)}$ can be determined using differentiation to be the regression coefficient from the regression of $\hat{y}_\alpha(n)$ on $\hat{c}_\alpha(n)$;

$$\theta = \frac{\sigma_{\hat{y}_\alpha(n),\hat{c}_\alpha(n)}}{\sigma^2_{\hat{c}_\alpha(n)}} = \frac{\sigma_{\hat{y}_\alpha(n)}}{\sigma_{\hat{c}_\alpha(n)}} \rho(\hat{y}_\alpha(n), \hat{c}_\alpha(n)) \tag{10}$$

where $\sigma_{\hat{y}_\alpha(n),\hat{c}_\alpha(n)}$ is the covariance of $\hat{y}_\alpha(n)$ and $\hat{c}_\alpha(n)$ and $\rho(\hat{y}_\alpha(n), \hat{c}_\alpha(n))$ is the correlation between $\hat{y}_\alpha(n)$ and $\hat{c}_\alpha(n)$.

3.1.2 Multiple Linear Controls

One can use multiple controls for variance reduction where $\hat{c}_\alpha(n)$ and θ become p-dimensional column vectors, $\underline{\hat{c}}_\alpha(n)$ and $\underline{\theta}$ with components $\hat{c}_{\alpha,i}(n)$ and θ_i, for $i = 1, \ldots, p$. With multiple controls, equation (9) becomes

$$\hat{y}'_\alpha(n) = \hat{y}_\alpha(n) - \underline{\theta}^T \left\{ \underline{\hat{c}}_\alpha(n) - \mathrm{E}[\underline{\hat{c}}_\alpha(n)] \right\}. \tag{11}$$

It can be shown (see Kendall and Stuart, 1977, chap. 27) that in the multiple control case, the values for $\underline{\theta}$ which minimize $\sigma^2_{\hat{y}'_\alpha(n)}$ are the multiple regression coefficients

$$\underline{\theta} = \left(\Sigma^{-1}_{\underline{\hat{c}}_\alpha(n)} \right) \sigma_{\hat{y}_\alpha(n),\underline{\hat{c}}_\alpha(n)} \tag{12}$$

where $\Sigma_{\underline{\hat{c}}_\alpha(n)}$ is the covariance matrix of $\underline{\hat{c}}_\alpha(n)$ and $\sigma_{\hat{y}_\alpha(n),\underline{\hat{c}}_\alpha(n)}$ is the p-dimensional vector with components $\mathrm{cov}(\hat{y}_\alpha(n), \hat{c}_{\alpha,i}(n))$, for $i = 1, \ldots, p$.

Rubinstein and Marcus (1985) demonstrated that the solution for $\underline{\theta}$ in the linear control of a single response, $\hat{y}_\alpha(n)$, is a special case of determining the canonical correlation coefficients for maximizing the correlation between linear combinations of multiple responses and multiple controls.

3.2 A Measure of the Effectiveness of a Control for Variance Reduction

One measure of effectiveness for a particular linear control is the percent variance reduction which involves the ratio of the variance of the controlled estimate $\hat{y}'_\alpha(n)$ to the uncontrolled estimate $\hat{y}_\alpha(n)$. A high percent variance reduction implies that the control is effective at reducing the variance of the point estimate. For a single control, assuming the optimal value for θ is known, the percent variance reduction is

$$1 - \frac{\sigma^2_{\hat{y}'_\alpha(n)}}{\sigma^2_{\hat{y}_\alpha(n)}} = \rho^2\left(\hat{y}_\alpha(n), \hat{c}_\alpha(n)\right). \tag{13}$$

Equation (13) implies that for the control to be effective, one should choose a random variable which is "strongly" correlated with $\hat{y}_\alpha(n)$ to be the control variable $\hat{c}_\alpha(n)$.

For multiple controls, the percent variance reduction is the direct generalization

$$1 - \frac{\sigma^2_{\hat{y}'_\alpha(n)}}{\sigma^2_{\hat{y}_\alpha(n)}} = R^2_{\hat{y}_\alpha(n),\underline{\hat{c}}_\alpha(n)}. \tag{14}$$

where

$$R^2_{\hat{y}_\alpha(n),\underline{\hat{c}}_\alpha(n)} = \frac{\sigma^T_{\hat{y}_\alpha(n),\underline{\hat{c}}_\alpha(n)}\left(\Sigma^{-1}_{\underline{\hat{c}}_\alpha(n)}\right)\sigma_{\hat{y}_\alpha(n),\underline{\hat{c}}_\alpha(n)}}{\sigma^2_{\hat{y}_\alpha(n)}}$$

is the square of the multiple correlation coefficient between $\hat{y}_\alpha(n)$ and $\underline{\hat{c}}_\alpha(n)$. As before, the effectiveness of the control depends upon a large value for $R^2_{\hat{y}_\alpha(n),\underline{\hat{c}}_\alpha(n)}$. When the number of multiple controls to use is given, one should simply choose those controls which maximize the $R^2_{\hat{y}_\alpha(n),\underline{\hat{c}}_\alpha(n)}$. However, determining the number of multiple controls to use is a more difficult problem which is complicated by the necessity of estimating the coefficients in $\underline{\theta}$.

3.3 Use of the Asymptotic Expected Value as an Approximation for the Expected Value of the Control

When using a linear control for variance reduction, the expected value of the control is subtracted from the control variable in the control function as in (9) so that the control function will have a mean of zero. A mean-zero control function is desirable when controlling an unbiased estimator such as a sample mean so that the controlled estimate is also unbiased. However, expected values of quantile estimators are rarely known exactly. If the values of the density function of C and its derivative at c_α are known, the biased expected value of the quantile estimator from (4) can be subtracted in the control function so that the control function does not affect the first order bias in the controlled quantile estimate. If the expected value of the biased quantile estimator is not known, it can be approximated by the asymptotic

expected value of the estimator; i.e. the actual quantile value c_α. The value c_α will replace $\mathrm{E}[\hat{c}_\alpha(n)]$ in the control function in (9). While this causes the control function to have order $1/n$ bias, there is already order $1/n$ bias in the estimate being controlled, $\hat{y}_\alpha(n)$, so that the order of the bias in the controlled estimate is the same as in the uncontrolled estimate.

Even when the biased expected value for the control from (4) is known, it may desirable to use the asymptotic value. There is empirical evidence, and it can be shown analytically, that use of a control function with order $1/n$ bias can actually decrease the magnitude of the first-order bias in the controlled estimate. For example, let $\mathcal{B}_{\hat{y}_\alpha(n)}$ denote the first order bias of $\hat{y}_\alpha(n)$ computed using (4) as $\mathcal{B}_{\hat{y}_\alpha(n)} = \mathrm{E}[\hat{y}_\alpha(n)] - y_\alpha + \mathrm{O}(1/n^2)$ and let $\mathcal{B}_{\hat{c}_\alpha(n)}$ denote the bias of $\hat{c}_\alpha(n)$ computed similarly. If using the linear control scheme (9) to control a quantile estimate, where $\mathcal{B}_{\hat{y}_\alpha(n)}/\mathcal{B}_{\hat{c}_\alpha(n)}$ is positive and

$$0 < \theta < 2\frac{\mathcal{B}_{\hat{y}_\alpha(n)}}{\mathcal{B}_{\hat{c}_\alpha(n)}},$$

the magnitude of the first-order bias of the controlled estimate is less than the magnitude of the first-order bias of the uncontrolled estimate.

If we are using sectioning to generate the overall point estimate and an estimate of the variance (standard deviation) of the point estimate, and we assume that θ is known, equations (6) and (7) can be combined with the linear control equation, (9), to get

$$\overline{\hat{y}'_\alpha}(m,n) = \frac{1}{m}\sum_{j=1}^{m}\hat{y}'_{\alpha,j}(n) \tag{15}$$

$$= \frac{1}{m}\sum_{j=1}^{m}\{\hat{y}_{\alpha,j}(n) - \theta\,(\hat{c}_{\alpha,j}(n) - c_\alpha)\} \tag{16}$$

with an unbiased estimate of the variance of the controlled estimate of

$$S^2_{\overline{\hat{y}'_\alpha}(m,n)} = \frac{1}{m(m-1)}\sum_{j=1}^{m}\left\{\hat{y}'_{\alpha,j}(n) - \overline{\hat{y}'_\alpha}(m,n)\right\}^2. \tag{17}$$

These results are straightforward. It is when θ is not known, the usual case, and has to be estimated using sectioning, that estimating the variance of the controlled estimate requires some care.

3.4 Estimating the Coefficients

In the usual case in simulation, the values for θ or $\underline{\theta}$ must be estimated since not enough information is known about the joint distribution of $\hat{y}_\alpha(n)$ and $\underline{\hat{c}}_\alpha(n)$ to determine the regression coefficients. For notation's sake, assume that one is using a single control. If using sectioning to estimate the point estimate along with its variance, the sectioned estimates $\hat{y}_j(n)$ and $\hat{c}_j(n)$, for $j = 1,\ldots,m$ are available to use to estimate θ. One could generate sample estimates of the variance and covariances in (10) to estimate θ; however since θ is the coefficient of regression, an equivalent but computationally more convenient method for estimating θ is to use linear least-squares regression.

The regression coefficient θ can be estimated by the least squares regression of $[\hat{y}_{\alpha,j}(n) - \overline{\hat{y}_\alpha}(m,n)]$ on $\theta[\hat{c}_{\alpha,j}(n) - c_\alpha]$ using the regression model

$$[\hat{y}_{\alpha,j}(n) - \overline{\hat{y}_\alpha}(m,n)] = \theta[\hat{c}_{\alpha,j}(n) - c_\alpha] + \epsilon_j, \quad j = 1, \ldots, m \tag{18}$$

where the $\hat{c}_{\alpha,j}(n)$ are considered fixed and ϵ_j is a mean-zero random variable independent of $\hat{c}_{\alpha,j}(n)$. Denote by $\widehat{\theta}(m,n)$ the estimate of θ from a regression which used m estimates for both the dependent variable and the predictor variable, where each of the estimates was based on n independent samples of Y or C as appropriate.

Once $\widehat{\theta}(m,n)$ is computed, the controlled estimate for each section can be computed using (9) as

$$\hat{y}'_{\alpha,j}(n) = \hat{y}_{\alpha,j}(n) - \widehat{\theta}(m,n)\,\{\hat{c}_{\alpha,j}(n) - c_\alpha\}\,. \tag{19}$$

where c_α is the approximation for the expected value of the control. The final controlled section estimate, $\overline{\hat{y}'_\alpha}(m,n)$, can be computed using (15) as the sample mean of the controlled estimates from each section. Unfortunately, estimating the variance of the $\overline{\hat{y}'_\alpha}(m,n)$ with (17) is not as straightforward since the individual $\hat{y}'_{\alpha,j}(n)$ are generally no longer independent because of the common $\widehat{\theta}(m,n)$. The characteristics of the quantile estimates and the variance estimates depend upon the joint distribution of $\hat{y}_\alpha(n)$ and $\hat{c}_\alpha(n)$.

3.4.1 Subtleties with the Joint Distribution of the Estimators

A key point of linear controls for quantile estimates is that the joint distribution of the statistic being controlled and *the control statistic*, here $\hat{y}_\alpha(n)$ and $\hat{c}_\alpha(n)$, is of primary importance for determining θ and the characteristics of the controlled estimate, *not* the joint distribution of the underlying populations Y and C.

This is in contrast to the use of a linear control for controlling an estimate of the mean, $\bar{y}$, with the sample mean of the control, $\bar{c}$. In this case, one can determine θ as a function of the joint distribution of Y and C since, using (10),

$$\theta = \frac{\mathrm{cov}(\bar{y},\bar{c})}{\mathrm{var}[\bar{c}]} = \frac{\mathrm{cov}(y,c)}{\mathrm{var}[c]}.$$

Although the joint distribution of $\bar{y}$ and $\bar{c}$ is different from the joint distribution of Y and C, one can estimate θ using estimates of the population covariances based on the N individual samples. In general, when controlling estimators other than the sample mean, one must estimate the covariances from the joint distribution of the controlled statistic and the control, not the joint distribution of the underlying populations.

3.4.2 Sectioning with the Assumption that the Joint Distribution is Multivariate Normal

If the joint distribution of $\hat{y}_\alpha(n)$ and $\hat{c}_\alpha(n)$ is multivariate normal and $\underline{\theta}$ is estimated, the point estimate of the quantile and the estimate of the variance of the point estimate have several nice properties:

- the controlled estimates for each section, $\hat{y}'_{\alpha,j}(n)$, are i.i.d. since the sample covariance matrix of the $\hat{c}_{\alpha,j}(n)$ is independent of their sample mean.
- $S^2_{\overline{\hat{y}'_\alpha}(m,n)}$, the estimate of the variance of $\overline{\hat{y}'_\alpha}(m,n)$ from (17) where $\hat{y}'_{\alpha,j}(n)$ is computed using (19), is an unbiased estimator, and
- one can develop an unconditional confidence interval for $\overline{\hat{y}'_\alpha}(m,n)$ using the t statistic following Lavenberg, Moeller and Welch (1982) since conditionally unbiased estimators remain unbiased unconditionally and conditional confidence intervals remain valid unconditionally (see Kendall and Stuart, 1977, p. 379).

When the multivariate normal assumption is not valid,

- the controlled estimates from each section $\hat{y}'_{\alpha,j}(n)$ are no longer independent since the sample mean and covariance matrix are no longer independent. The controlled estimates also have additional $O(1/m)$ bias from the estimation of θ.
- $S^2_{\overline{\hat{y}'_\alpha}(m,n)}$ from (17) can still be used to estimate the variance of $\overline{\hat{y}'_\alpha}(m,n)$ although it is now biased, and
- even if the $\hat{y}'_{\alpha,j}(n)$ are normally distributed, a confidence interval based on a t statistic is only approximate because of the lack of independence of the individual section estimates.

One method for maintaining independence between the controlled section estimates at the cost of a loss of variance reduction is to estimate θ independently for each section.

3.4.3 Subsectioning

An alternative to estimating a single $\widehat{\theta}(m,n)$, which couples the $\hat{y}'_{\alpha,j}(n)$ together so that they are no longer independent, is to generate an individual estimate of θ for each section. This can be done by subsectioning the n samples within the section and calculating quantile estimates within the section to use as data to estimate $\widehat{\theta}_j(v,l)$. More formally, for each jth section, for $j = 1,\ldots,m$,

1. divide the n samples into v subsections of length l where $v \times l = n$, and
2. estimate $\hat{y}_{\alpha,j,k}(l)$ and $\hat{c}_{\alpha,j,k}(l)$ for each kth subsection, for $k = 1,\ldots,v$.
3. Use the v sets of subsection estimates $\hat{y}_{\alpha,j,k}(l)$ and $\hat{c}_{\alpha,j,k}(l)$ from the jth section to estimate $\widehat{\theta}_j(v,l)$ using a regression model similar to (18).

Once $\widehat{\theta}_j(v,l)$ has been estimated, the controlled estimate for the jth section is computed as

$$\hat{y}'_{\alpha,j}(n) = \hat{y}_{\alpha,j} - \widehat{\theta}_j(v,l)\,(\hat{c}_{\alpha,j}(n) - c_\alpha)\,. \tag{20}$$

The equation is similar to (19) only now there is a subscript on $\widehat{\theta}$, which also has different arguments. The final controlled estimate is calculated as before, as a sample

mean using (15), and the estimate of variance of the point estimates is calculated using (17).

An advantage of subsectioning is that by using an independent estimate of θ to calculate each section's controlled estimate, the $\hat{y}'_{\alpha,j}(n)$ are now i.i.d.. A disadvantage of using subsectioning is the loss of predicted variance reduction. This occurs for two reasons. The first is that instead of needing one estimate of θ, now m estimates are needed and each additional estimate tends to reduce the achieved percent variance reduction. The second reason is that $\widehat{\theta}(v,l)$ is not an unbiased estimator of the regression coefficient for $\hat{y}_\alpha(n)$ and $\hat{c}_\alpha(n)$ since it is calculated using quantile estimates based on l samples, which have a different joint distribution than $\hat{y}_\alpha(n)$ and $\hat{c}_\alpha(n)$. There can also be some additional bias in the $\hat{y}'_{\alpha,j}(n)$ from the estimation of θ_j.

3.4.4 Splitting and The Jackknife

Other methods which have been used with linear controls for calculating a point estimate and the variance of the point estimate include splitting and the jackknife. Each of these techniques is described in Lewis and Orav (1989, chap. 9) and in Nelson (1988).

The splitting technique removes the bias caused by estimating θ with the same data being controlled at the cost of reducing the percent variance reduction. Splitting has been described in Tocher (1963, p. 115) and then in Beale (1985). When using sectioning to generate m individual section quantile estimates $\hat{y}_{\alpha,j}(n)$ and $\hat{c}_{\alpha,j}(n)$, for $j = 1,\ldots,m$, the splitting procedure generates an estimate of θ for each section. The estimate of θ for the jth section is computed using all of the section estimates except the jth set of estimates. The controlled estimate for each section is computed using (20) with $\widehat{\theta}_j(m-1,n)$. The final controlled estimate and its variance are computed as before as the sample mean of the individual controlled section estimates and the sample variance of the sample mean

The splitting estimator eliminates the bias in $\hat{y}'_{\alpha,j}(n)$ due to estimating θ. However, like the sectioning estimator it has the disadvantage that the $\hat{y}'_{\alpha,j}(n)$ are no longer independent. It also has the same disadvantage as the subsection estimator in that m estimates of θ must be computed, reducing the percent variance reduction. The primary purpose for using the splitting estimator has been to eliminate the $O(1/m)$ bias in the controlled estimate from the estimation of θ in non-normal samples when controlling unbiased estimators. Since the quantile estimator already has $O(1/n)$ bias, which is unaffected by splitting, and splitting has no other clear advantages over the section or subsection estimator, we chose not to use it.

Jackknifing is a method for removing the $O(1/n)$ bias in $\hat{y}_\alpha(n)$ at the price of uncertainty about the loss of percent of variance reduction in small to medium sized samples. For an "m-fold"jackknife estimate, one combines an estimate based on the entire data set, $\hat{y}_{\alpha,0}(N)$, with m estimates, each based on the data set with N/m samples deleted, $\hat{y}_{\alpha,j}(N-m)$, for $j = 1,\ldots,m$, to get a set of m 'pseudo values" ${}_{(j)}\hat{y}_\alpha(N-m)$, for $j = 1,\ldots,m$. The final jackknife point estimate is the sample mean of the pseudo values. In some circumstances, one can also use the sample variance of the sample mean of the pseudo values as an estimate of the variance of the jackknife

point estimate.

The jackknife estimate has an advantage over the section and subsection estimators in that the bias of the quantile estimates is reduced since each pseudo value is based on estimates using $N - m$ instead of N/m samples. Unfortunately it has some disadvantages as well. Lavenberg, Moeller and Welch (1982) examined the use of the jackknife when using a linear control for the sample mean under the assumption of a multivariate normal distribution between the statistic of interest and the control. They found that the jackknifed confidence interval was usually larger and more computationally expensive than the standard linear control based confidence interval. Nelson (1988) compared the performance of several methods for linear control of the mean when the normality assumption was violated and found that the jackknife was usually "dominated" by the splitting estimator.

The jackknife has been used in quantile estimation. Seila (1982) used a 2-fold jackknife for removing the bias of quantile estimates however he used a sectioning approach for estimating the variance of the point estimate, not the jackknife estimate for the variance of the point estimate. Miller (1974), and Efron and Gong (1983) imply that the jackknife technique may not be an appropriate tool for use with quantile estimation because of the discontinuous, nonlinear nature of quantile estimators such as (2). Our empirical results (presented in the last section) confirmed that the jackknife was not suitable for computing quantile estimates and estimates of the variance of the jackknife point estimate because of the high variability of the point estimates and the poor performance of the jackknife estimate of the variance of the jackknife point estimator.

3.5 The Loss Factor

In general, regardless of the method chosen, estimating the coefficients can cause a reduction in the percent variance reduction predicted by (13) or (14). Lavenberg, Moeller and Welch (1982) investigated the decrease in predicted variance reduction caused by using the individual samples to estimate $\underline{\theta}$ for a linear control of the sample mean. Under the assumption of multivariate normality between the statistic of interest and the control, they concluded that the decrease in variance reduction due to estimating $\underline{\theta}$ could be predicted by multiplying the $R^2(\cdot)$ in (14) by a "loss factor". The loss factor was $(m-2)/(m-p-2)$ where m was the number of independent samples of the statistic being controlled and p was the number of controls whose coefficients had to be estimated. The loss factor is a deterrent to adding more controls simply to achieve a small increase in the R^2 in (14). As one selects more controls for a multiple control scheme, the impact of the loss factor can quickly overcome the benefits of increasing the R^2. Thus one can not guarantee an improvement in the effectiveness of a linear control by simply adding more controls.

3.6 Measuring the Effectiveness of a Control at Reducing Sample Sizes

Lewis and Orav (1989, p. 262) mention an alternative measure for quantifying the effectiveness of a control scheme. They look at the square root of the ratio of the variance of the uncontrolled estimate to the variance of the controlled estimate. This ratio can be considered to be the ratio of the sample size that would be needed

to achieve a given standard deviation without using the control scheme, to the sample size needed to achieve the same standard deviation using the control. When expressed in terms of the correlation coefficient for the controlled statistic and the control, the ratio becomes $1/(1-\rho^2(\cdot))^{1/2}$. Given a value for $\rho(\cdot)$, the formula gives the increase in the sample size that would be needed to achieve the same standard deviation without the control. Given a desired reduction in sample size, say 1/2, the formula implies that to achieve a given standard deviation while cutting the sample size in half, one must have $1-\rho^2 = .25$, which implies a correlation coefficient of ± 0.86.

Linear controls are typically unable to reduce the sample size by as much as a half because the correlation between the statistic of interest and a linear function of the control variables is not high enough. Since many statistics have a nonlinear relationship with the control variables, one possible means for increasing the variance reduction for a given set of controls is to allow nonlinear transformations of the controls.

4 NONLINEAR CONTROLS

4.1 Definition of a Nonlinear Control

One can generalize the linear control scheme for p controls, (11), to include nonlinear transformations of random variables as controls for variance reduction as shown in Lewis, Ressler and Wood (1989). Let $h_i(\hat{c}_{\alpha,i}(n), \underline{\theta}_i)$, for $i = 1, \ldots, p$, be a transformation function of the random variable $\hat{c}_{\alpha,i}(n)$ and let $\underline{\theta}_i$ be a vector of coefficients where, depending upon $h_i(\cdot)$, the vector $\underline{\theta}_i$ may have more than one component. When incorporating nonlinear transformations of multiple controls, the linear control scheme (11) becomes

$$\hat{y}'_\alpha(n) = \hat{y}_\alpha(n) - H(\underline{\hat{c}}_\alpha(n), \underline{\theta}) \tag{21}$$

where for our purposes $H(\cdot)$ is a linear additive combination of the p transformed controls, $h_i(\hat{y}_{\alpha,i}(n), \underline{\theta}_i)$, and their expected values, $\mathrm{E}[h_i(\hat{y}_{\alpha,i}(n), \underline{\theta}_i)]$, for $i = 1, \ldots, p$. The vector $\underline{\theta}$ contains the coefficients from the linear combination in addition to the p sets of coefficients from the individual transformations. $H(\underline{\hat{c}}_\alpha(n), \underline{\theta})$ will be referred to as the control function. A control function with terms that are nonlinear in the unknown coefficients will be said to be a nonlinear control. For ease of notation, the coefficients $\underline{\theta}$ may be suppressed in the expressions for $H(\cdot)$ and $h(\cdot)$. When there is only one control so that $p = 1$, the subscript i will be suppressed so that $h_i(\cdot) = h(\cdot)$.

In some simulations possible control variables may have very low correlation with $\hat{y}_\alpha(n)$. For a given control, two of the possible sources for the low correlation between $\hat{y}_\alpha(n)$ and $\hat{c}_\alpha(n)$ are:

1. there is in fact very little structural relationship between $\hat{y}_\alpha(n)$ and the control; i.e. a bivariate scatter plot of $\hat{y}_\alpha(n)$ versus $\hat{c}_\alpha(n)$ would look patternless, or

2. the structural relationship between $\hat{y}_\alpha(n)$ and $\hat{c}_\alpha(n)$ is of a nonlinear form which is poorly approximated by a straight line.

In the first case, a nonlinear control may or may not offer improvement over the linear control. In the second case, a nonlinear control can offer substantial improvement in variance reduction, as shown in Lewis, Ressler and Wood (1989).

A simple example will show the potential benefits of nonlinear transformations. Let z be a Normal (0,1) random variable which is being used to control the sample mean of $w = z^2$. It follows that

$$\text{cov}(w, z) = \text{E}[z^3] - \text{E}[z^2]\text{E}[z] = 0$$

so that $\rho(w, z)$ is zero, which implies zero effectiveness for the linear control as well. Now allow the nonlinear transformation

$$h^*(z) = h(z, \theta) = z^\theta$$

with $\theta = 2$. The transformed random variable $h^*(z)$ is a χ^2_1 random variable with mean 1 and variance 2. It follows that

$$\text{cov}(w, h^*(z)) = \text{var}[z^2] = 2 \Longrightarrow \rho(w, h^*(z)) = \frac{2}{2} = 1$$

so that the nonlinear control is completely effective. Therefore when evaluating a potential control, one should ask: *Can this random variable be transformed to have a "high" correlation with the statistic of interest?*

4.2 The Existence of Optimal Nonlinear Transformations

For some random variables, transformations do exist which will improve their correlation with $\hat{y}_\alpha(n)$.

- Let $\hat{y}_\alpha(n)$ and $\underline{\hat{c}}_\alpha(n)$, with p components $\hat{c}_{\alpha,i}(n)$, for $i = 1, \ldots, p$, be random variables with a general but nonsingular joint distribution.
- Let $g(\hat{y}_\alpha(n)) = g(\hat{y}_\alpha(n), \phi)$ and $h_i(\hat{c}_{\alpha,i}(n)) = h(\hat{c}_{\alpha,i}(n), \underline{\theta}_i)$, for $i = 1, \ldots, p$ be mean-zero transformation functions of random variables $\hat{y}_\alpha(n)$ and $\hat{c}_{\alpha,i}(n)$ such that $\text{var}[g(\hat{y}_\alpha(n))] = 1$ and $\text{var}[h_i(\hat{c}_{\alpha,i}(n))] < \infty$, for $i = 1, \ldots, p$.

Breiman and Friedman (1985) proved the existence of optimal transformations for maximizing the correlation between $g(\hat{y}_\alpha(n))$ and $H(\underline{\hat{c}}_\alpha(n))$, a linear additive function of the mean-zero $h_i(\hat{c}_{\alpha,i}(n))$. The optimal transformation for one variable can be expressed in terms of the conditional expected values of given transformations of the other variables. In the bivariate case, where $H(\cdot) = h(\cdot)$ since $p = 1$, the pair of optimal transformations $g^*(\cdot)$ and $h^*(\cdot)$ are:

$$g^*(\hat{y}_\alpha(n)) = \frac{\text{E}[h^*(\hat{c}_\alpha(n)) \mid \hat{y}_\alpha(n)]}{\| \text{E}[h^*(\hat{c}_\alpha(n)) \mid \hat{y}_\alpha(n)] \|}$$

and

$$h^*(\hat{c}_\alpha(n)) = \text{E}[g^*(\hat{y}_\alpha(n)) \mid \hat{c}_\alpha(n)]$$

where $\| \cdot \| = \{\text{E}[(\cdot)^2]\}^{1/2}$.

In the multiple control case, where $p > 1$,

$$g^*(\hat{y}_\alpha(n)) = \frac{E\left[\sum_{i=1}^{p} h_i^*(\hat{c}_{\alpha,i}(n)) \mid \hat{y}_\alpha(n)\right]}{\left\| E\left[\sum_{i=1}^{p} h_i^*(\hat{c}_{\alpha,i}(n)) \mid \hat{y}_\alpha(n)\right] \right\|} \tag{22}$$

and

$$h_i^*(\hat{c}_{\alpha,i}(n)) = E\left[g(\hat{y}_\alpha(n)) - \sum_{j \neq i} h_j^*(\hat{c}_{\alpha,j}(n))\right]. \tag{23}$$

The transformations $g^*(\cdot)$ and $h^*(\cdot)$ in (22) and (23) will usually be nonlinear, the exception being when $\hat{y}_\alpha(n)$ and $\hat{\underline{c}}_\alpha(n)$ have a multivariate normal distribution.

Results from Lancaster (1966) can be used to show that if $\hat{y}_\alpha(n)$ and $\hat{\underline{c}}_\alpha(n)$ have a multivariate normal distribution, the solutions for $g(\hat{y}_\alpha(n))$ and $H(\hat{\underline{c}}_\alpha(n))$ which have maximal correlation between $g(\hat{y}_\alpha(n))$ and $H(\hat{\underline{c}}_\alpha(n))$, over all measurable functions of finite variance, are the linear transformations which yield the first Hotelling canonical variables. In other words, when $\hat{y}_\alpha(n)$ and $\hat{\underline{c}}_\alpha(n)$ have a multivariate normal distribution, using the linear control scheme (11), with the multiple regression coefficients for $\underline{\ell}$, produces the greatest amount of variance reduction. Conversely, whenever the joint distribution of $\hat{y}_\alpha(n)$ and $\hat{\underline{c}}_\alpha(n)$ is not multivariate normal, a nonlinear control offers the possibility for greater variance reduction over a linear control.

4.3 Estimating the Optimal Nonlinear Transformations

Determining the optimal transformations in (22) and (23) analytically requires the joint distribution of $\hat{y}_\alpha(n)$ and $\hat{\underline{c}}_\alpha(n)$ which, in the context of a simulation, is unknown. In the multivariate normal case, the form of the transformations are known to be linear and one can estimate the coefficients using one of the methods described earlier. With a nonlinear control, one must first estimate the form of the transformations.

Breiman and Friedman (1985) also developed the Alternating Conditional Expectation Algorithm (ACE) as a means for generating nonparametric estimates of the optimal transformations (22) and (23). In the ACE implementation for finite data sets of continuous variables, data smooths are used in place of the analytical conditional expected values. The ACE algorithm produces estimates of the optimal transformations as sets of fitted values, one set for each variable. Plotting the fitted values against the original values gives the shape of the estimated transformation for each variable. ACE also provides an estimate of the maximum obtainable squared correlation between the transformed response and the sum of transformed predictors. This R^2 estimate is useful as it provides an estimate of an upper bound on the percent variance reduction one can obtain using the given set of controls.

Since ACE does not give an explicit analytical form for its estimate of the optimal transformation, one must approximate the optimal transformation with a parametric nonlinear transformation. The output from ACE is useful in selecting an appropriate

approximating transformation. One possible approximating transformation is the scaled power transformation

$$h(\hat{c}_\alpha(n), \theta) = \frac{(\hat{c}_\alpha^\theta(n) - 1)}{\theta}, \quad \text{for } \theta > -1, \tag{24}$$

where θ is an unknown parameter which becomes a coefficient which must be estimated. Using this transformation, the nonlinear control scheme (21) can become

$$\hat{y}'_\alpha(n) = \hat{y}_\alpha(n) - \theta_1 \left\{ \frac{\hat{c}_\alpha^{\theta_2}(n) - 1}{\theta_2} - \mathrm{E}\left[\frac{\hat{c}_\alpha^{\theta_2}(n) - 1}{\theta_2} \right] \right\} \tag{25}$$

where both θ_1 and θ_2 need to be estimated. Other possible transformations are described in Lewis, Ressler and Wood (1989).

As a general rule, a transformation should contain the linear transformation as a special set of parameter values $\underline{\theta}_L$. This allows for the linear control to be a special case of the nonlinear control when the joint distribution between the statistic of interest and the control is multivariate normal. Choosing the special set of parameter values $\underline{\theta}_L$ as starting values for the nonlinear optimizer which estimates the coefficients initializes the optimizer at the linear control. Any movement made by the optimizer away from the starting values implies that the nonlinear control is giving improved variance reduction over the linear control. Thus using a nonlinear control, one can not do worse than using a linear control.

One of the problems in choosing an approximating transformation $h_i(\hat{c}_{\alpha,i}(n), \underline{\theta})$ is that $\mathrm{E}[h_i(\hat{c}_{\alpha,i}(n), \underline{\theta})]$ must be known exactly or approximately. This severely limits the selection of nonlinear transformations available to approximate $h_i^*(\hat{c}_{\alpha,i}(n))$ as the necessary expected values may be intractable or unknown for some transformations. The difficulty in analytically determining the expected value of the transformed control can be greatly reduced when using monotone transformations of quantile estimators as controls, as is discussed in the next section.

5 NONLINEAR CONTROL OF QUANTILE ESTIMATES

5.1 The Behavior of Quantiles Under Monotone Transformations

Quantiles have a property that is especially useful when working with nonlinear controls. Under strictly monotone transformations of the underlying random variable, the quantiles transform monotonely as well. For example,

- let $h(\cdot)$ be a strictly monotone function with inverse $h^{-1}(\cdot)$,
- let C be a random variable with a continuous, strictly monotone cumulative distribution function such that for all α between zero and one, $F_C^{-1}(\alpha) = c_\alpha$, and
- let $W = h(C)$ be the transformed random variable.

By definition of a quantile,

$$\Pr\{C \leq c_\alpha\} = \alpha \text{ and } \Pr\{W \leq w_\alpha\} = \alpha.$$

Therefore:

$$\begin{aligned} \Pr\{W \leq w_\alpha\} &= \Pr\{h(C) \leq w_\alpha\} \\ &= \Pr\left\{C \leq h^{-1}(w_\alpha)\right\} = \alpha. \end{aligned}$$

This implies that for all α between zero and one,

$$w_\alpha = h(c_\alpha). \tag{26}$$

For example, if C has a uniform (0,1) distribution with .9 quantile of $c_{.9} = .9$, then the .9 quantile of $W = h(C) = C^2$, namely $w_{.9}$ is equal to $c_{.9}^2 = .9^2 = .81$.

The key point is that the α quantile of a transformed random variable can be found by applying the same transformation to the α quantile of the original random variable.

5.2 Controlling Quantile Estimates

The fact that quantiles transform monotonely under strictly monotone transformations of the underlying random variable can also be useful in computing the expected value of a transformed quantile estimator. It is important to note that the random variable being transformed is the quantile estimator $\hat{c}_\alpha(n)$ and not the underlying C. For a given nonlinear transformation, it may be possible to compute the expected value of $h(\hat{c}_\alpha(n))$. For example, if C has a uniform (0,1) distribution, and $h(\hat{c}_\alpha(n))$ is the scaled power transformation, (24) where θ is constrained to be non-negative, $h(\hat{c}_\alpha(n))$ has a Beta distribution with a known expected value. For other distributions of $\hat{c}_\alpha(n)$, or other transformations $h(\cdot)$, the expected value may not be tractable. This is where the use of strictly monotone transformations can help.

We are interested in the expected value of the transformed quantile estimator. When a strictly monotone transformation is applied to the underlying C, the quantile estimator $\hat{c}_\alpha(n)$ transforms monotonely as well, i.e. if $\hat{c}_\alpha(n)$ estimates c_α and $h(C) = W$, with α quantile w_α, then

$$\hat{w}_\alpha(n) = h(\hat{c}_\alpha(n)).$$

From the point of view of the quantile estimator, applying a strictly monotone transformation to a quantile estimator, $\hat{c}_\alpha(n)$, yields the same estimate as using the identical transformation on the underlying random variable C and then using (2) to estimate the α quantile. Although for small n

$$E[h(\hat{c}_\alpha(n))] \neq h(E[\hat{c}_\alpha(n)]),$$

it is true that as $n \longrightarrow \infty$,

$$E[h(\hat{c}_\alpha(n))] \longrightarrow h(c_\alpha) \text{ and } h(E[\hat{c}_\alpha(n)]) \longrightarrow h(c_\alpha)$$

so that asymptotically, the expected value of the transformed quantile estimator is the same as the expected value of the quantile estimator of the transformed underlying random variable.

Since the asymptotic expected values are the same, if the individual transformation functions $h(\cdot)$ in the control function $H(\hat{c}_\alpha(n), \underline{\theta})$ are restricted to strictly monotone transformations, one can approximate $\mathrm{E}[h(\hat{c}_\alpha(n), \underline{\theta})]$ in the nonlinear control function $H(\hat{c}_\alpha(n), \underline{\theta})$, with the asymptotic expected value of the transformed control, namely, the transformed value of the α quantile, $h(c_\alpha, \underline{\theta})$. Calculating $h(c_\alpha, \underline{\theta})$ is trivial since c_α is a constant. Using the asymptotic expected value with the scaled power transformation, the nonlinear control scheme becomes

$$\hat{y}'_\alpha(n) = \hat{y}_\alpha(n) - \theta_1 \left\{ \frac{\hat{c}_\alpha(n)^{\theta_2} - 1}{\theta_2} - \frac{c_\alpha^{\theta_2} - 1}{\theta_2} \right\}.$$

The use of the approximation introduces bias into the control function, but it is still $\mathrm{O}(1/n)$ and may, as in the linear control case, reduce the magnitude of the first order bias of the controlled estimate. *The key point is that the analytical burden of calculating the expected value of the transformed control has been greatly reduced.*

Once the approximating transformations for the $\underline{\hat{c}}_\alpha$ have been selected, one can use either the section or subsection estimator to estimate $\underline{\theta}$ and calculate the final, controlled point estimate $\overline{\hat{y}'_\alpha}(m, n)$ in (15) and an estimate of the variance of the point estimate. Regardless of the method, the coefficients in $\underline{\theta}$ for $h(\underline{\hat{c}}_\alpha, \underline{\theta})$ can be estimated using a nonlinear least-squares regression algorithm as the nonlinear optimizer.

5.3 Selection of m and n for a Nonlinearly Controlled Section Estimate when $\underline{\theta}$ Must be Estimated

A major factor that must also be considered in the selection of m and n for fixed sample size N is the impact of n, the number of samples used to compute the individual quantile estimates, on the joint normality of the quantile estimates. When computing a controlled section estimate and estimating the coefficients $\underline{\theta}$, the impact of m and n on the variance of the estimate $\underline{\hat{\theta}}(m, n)$ must also be considered.

As previously discussed, given a fixed sample size N the values of m and n which minimize the mean square error of the crude section estimate are a function of the coefficients in the asymptotic expansions for the mean and variance of the estimator, equations (4) and (5). The variance of the controlled estimate $\hat{y}'_\alpha(n)$ is a function of the variance of the estimate of the coefficients $\underline{\theta}$ in addition to the variance of the crude estimate, $\hat{y}_\alpha(n)$, and the variance of the estimate of the control $\hat{c}_\alpha(n)$. In general, the bias and variance of coefficients estimated via least-squares nonlinear regression is a decreasing function of the number of estimates used as data in the regression (see Gallant, 1987, chap. 1). When using the section estimator, this implies that one would like m, the number of quantile estimates, to be large. However, as m increases for fixed N, n must decrease, increasing the bias and variance of the estimates used as data in the regression. If n is too small, the bias and variance of the estimates could be such that there is actually very

little nonlinear or even linear relationship between the crude and control quantile estimates so that any control scheme is ineffective.

If n, the number of samples in a section, is too large, the joint distribution of the crude and control quantile estimates approaches a joint normal distribution as seen in part 2.1. The impact of the joint normality is that the optimal nonlinear transformation is now the linear transformation of the linear control as seen in part 4.2 and one has lost the increased effectiveness of the nonlinear control. This result is similar to one obtained by Glynn and Whitt (1989) who state that "no improvement in asymptotic efficiency can be achieved by generalizing the notion of control variables from a linear form to a nonlinear setting." They go on to say however, "...this does not preclude the possibility of better performance by nonlinear methods in a small sample context." The key point is that by avoiding the asymptotic joint normality through keeping small the number of samples used to compute the individual quantile estimates, the nonlinear controls can be more effective than the asymptotic linear controls.

When using the subsection estimator, the interplay between m and n changes. One must now consider the impact of choices for v, the number of subsection estimates, and l, the number of samples used to compute a subsection estimate. With the section estimator one wanted m, as the number of points in the regression, to be large. For the subsection estimator m is the number of estimates of $\underline{\theta}$ to compute and a large m implies more regression computations that have to be made, as well as a small value for n. For any given value of n, the choice of v and l has slightly different considerations than the choice of m and n for the section estimator. An important consideration for the subsection estimator is that l be "close" to n so that the joint distribution $\hat{y}_\alpha(l)$ and $\hat{c}_\alpha(l)$ will be similar in shape to that of $\hat{y}_\alpha(n)$ and $\hat{c}_\alpha(n)$. If the two joint distributions are not similar in shape, then the subsection estimate of $\underline{\theta}$ could be very biased, reducing the effectiveness of the control. This suggests making v as small as possible while still being two to three times the number of coefficients being estimated. If n is too small, the few samples available for the v subsections of length l will force both v and l to be small, resulting in possibly little structure to exploit, or unreliable estimates of θ, both of which result in ineffective control. The solution would seem to be to make n large.

Making n too large results in the same problems for the subsection estimator as it did for the section estimator. If n is too large, there are few controlled section estimates which reduces the precision of the variance estimate. More importantly, n is still the critical factor for the joint normality of the estimate being controlled and the control estimate. If n is too large, the asymptotic joint normality reduces the effectiveness of the linear control to that of the linear control.

The selection of m and n for a fixed N which minimizes the bias, variance or mean square error of the controlled estimate is a complicated function of many parameters. These parameters include the value of α, the sample size N, and unfortunately, because of the need to estimate $\underline{\theta}$, characteristics of the unknown joint distribution of the underlying populations Y and C. An alternative to attempting to estimate the optimal m and n via a functional approximation is to use graphical methods to assist in the selection of m and n such as in Heidelberger and Lewis (1981). In the experiment described below, for a given fixed sample size N, the results of using

different values of n are compared graphically as well as numerically to assist in selecting m and n.

6 THE SIMULATION EXPERIMENT

6.1 The Factors

The simulation experiment used M replications to investigate simulation procedures for estimating the α quantile of a distribution and estimating the variance of the quantile estimate. The factors in the simulation experiment included the distribution of the underlying population of interest, the value for α, the method of estimating the quantile, the sample size, the choice of m and n for the section estimator and the choice of the m for the m-fold jackknife estimator. All of the computations were performed in the APL2-based statistcal computing package GRAFSTAT.

6.2 The Statistic of Interest

The distribution used in the results presented here was suggested by Hsu and Nelson (1987). The statistic of interest is the estimator for the α quantile of a random variable Y where

$$Y = \left(\frac{1}{1.01 - X}\right) 100 + \epsilon$$

and X has a uniform (0,1) distribution and ϵ has a uniform (0,.5) distribution and is independent of X. The untransformed control is the estimator of the α quantile of X. The value of α will be .95 for the results presented here. The true value for the .95 quantile of Y, namely $y_{.95}$, is .164167.

Figure 1 shows the nonlinear nature of the relationship between $\hat{y}_\alpha(n)$ and $\hat{x}_\alpha(n)$ for four values of n with the sample size N fixed at 1000. Prior to plotting, the quantile estimates were standardized by subtracting off the sample mean of the quantile estimates from each estimate, and then dividing each estimate by the sample standard deviation of the quantile estimates. Thus the "true" values are zero. The quantile estimates were standardized so that one could visually assess the correlation between the quantile estimator of interest and the control quantile estimator. Note that the scales of the axes in Figure 1 change as n increases to 100, 250 and 500 as the ranges of the standardized quantile estimates become more concentrated about the true values of zero.

For $n = 25$ in Figure 1, the relationship between $\hat{y}_\alpha(n)$ and $\hat{x}_\alpha(n)$ is highly nonlinear. As n increases to 100, 250 and 500 the relationship seems to become more linear as the number of estimates available decreases to just two at $n = 500$ where with only two pairs of estimates, the relationship must appear linear. However, one can see from Figure 2, where $N = 6000$, that even for $n = 1000$ the relationship between $\hat{y}_\alpha(n)$ and $\hat{x}_\alpha(n)$ still has nonlinear tendencies. In all cases, the relationship appears to be one that would be well approximated by a monotone transformation.

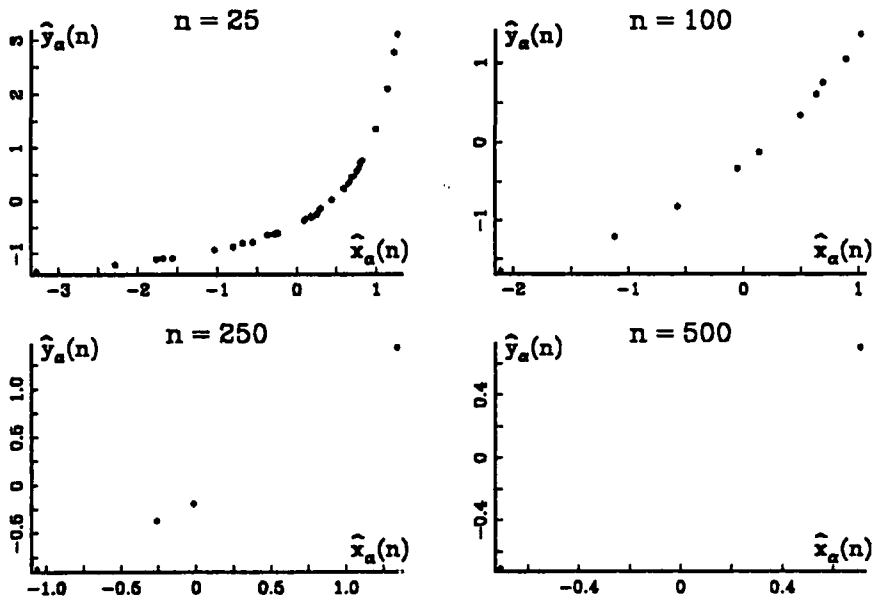

Figure 1: Scatterplots illustrating the joint distribution of standardized section point estimates of the .95 quantile of Y and X for n = 25, 100, 250, and 500 from a sample of N = 1000 samples. Since the estimates are standardized, the true values are zero.

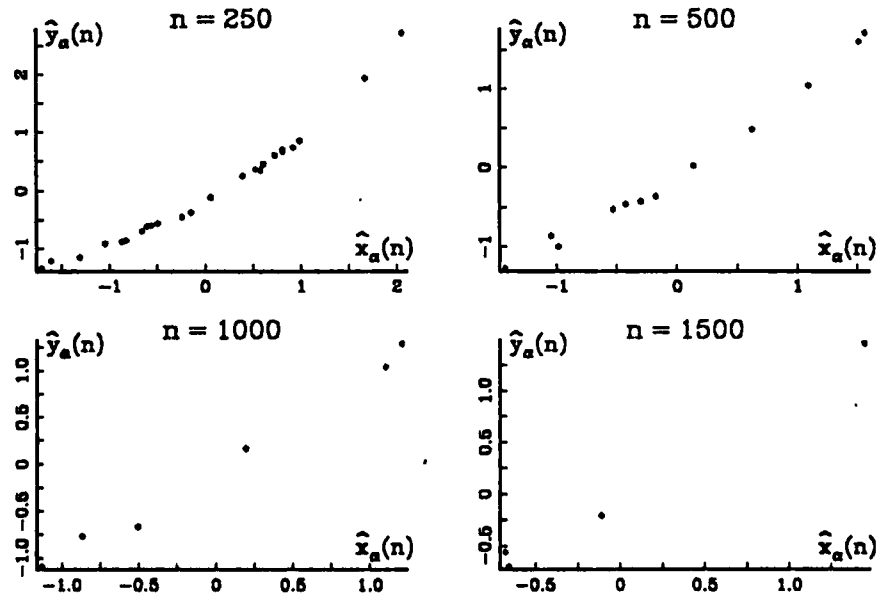

Figure 2: Scatterplots illustrating the joint distribution of of standardized section point estimates of the .95 quantile of Y and X for n = 250, 500, 1000, and 1500 from a sample of N = 6000 samples. Since the estimates are standardized, the true values are zero.

6.3 The Section Estimator versus the Jackknife Estimator

As stated previously, the section estimator was preferred over the jackknife estimator for estimating the α quantile along with an estimate of the variance (standard deviation) of the quantile estimator. Analytically, the section estimator of the variance of the section estimate from (17) is an unbiased estimator and the section estimate of the standard deviation has O(1/m) bias. We will graphically show the performance of the section estimate of the standard deviation so that the graphs can be compared with the performance of the jackknife estimation procedure.

The performance of the section estimator can be seen in Figure 3. The top graph of Figure 3 shows a series of boxplots of section point estimates of the .95 quantile of Y calculated using (6). For a discussion of boxplots see Chambers et. al. (1983, chap. 2). The boxplots summarize the distribution of the section estimates, for varying n, from 300 independent replications of $N = 1000$ samples. The data under the graph are the sample statistics from the 300 estimates in each boxplot. The bottom graph consists of boxplots of section estimates of the standard deviation, calculated using (7), corresponding to the point estimates in the top graph, again with the sample statistics underneath.

The top graph in Figure 3 shows that as n increases from 10 to 500, for a fixed sample size $N = 1000$, the bias in the section point estimates tends to decrease as expected. However, the top graph also shows that increasing n does not necessarily decrease the sample variance of the section quantile estimator because of the impact of decreasing the number of estimates, m, with which the section point estimate of the quantile is computed.

The bottom graph of Figure 3, of the section estimates of the standard deviation of the section point estimate, shows another effect of increasing n. As n increases and m decreases, it is easy to see that the standard deviation of the estimates of the standard deviation also increases, from .00227 for $n = 10$ to .01170 for $n = 500$, so that the section estimate of the standard deviation becomes less precise. As the section estimate of the standard deviation has $O(1/m)$ bias, one would expect that the section estimate of the standard deviation should be closer to the estimate of the sample standard deviation for small n. A check of the sample standard deviation in the top graph against the mean of the section estimates of the standard deviation in the bottom graph shows that in fact the two values of .02030 and .01974 are fairly close at $n = 10$ and become farther apart as n increases. The significance of the difference will be examined in a moment.

Figure 4 shows the performance of the jackknife estimator for y_α. The top boxplots are the m-fold jackknife estimate of the .95 quantile of Y, for varying m, from the same 300 independent replications of $N = 1000$ samples used for the section estimates in Figure 3. The data under the graph are the sample statistics from the 300 estimates in each boxplot. The bottom graph in Figure 4 consists of boxplots of the corresponding jackknife estimates of the standard deviation of the jackknife point estimates in the top graph, again with the sample statistics underneath.

The top graph in Figure 4 shows that for a fixed sample size $N = 1000$, the jackknife estimates become highly variable as m increases, as well as having in general a slight positive bias ($y_\alpha = .164167$). The main reason for not using the jackknife

technique however is the poor performance of the jackknife estimate of the standard deviation of the point estimate. A check of the sample standard deviation in the top graph against the mean of the jackknife estimates of the standard deviation in the bottom graph shows that the two estimates of the standard deviation become quite far apart as m increases. For $m = 2$ the values are the closest, at .02202 for the sample standard deviation of the point estimate and .01555 for the jackknife estimate of the standard deviation of the point estimate

The purpose of estimating the standard deviation of the point estimators is to have a measure of the precision of the point estimate. The section and jackknife estimators of the standard deviation of the point estimate are both trying to estimate the standard deviation of a sample of section or jackknife point estimates. To more formally assess their performance we used the data from the 300 independent replications previously shown in Figures 3 and 4. The procedure used for both the section estimates and the jackknife estimates was as follows:

1. The point estimates from the 300 replications were sectioned into 30 independent sections of 10 point estimates each. The sample standard deviation was computed for each of the 30 sections. Thus there were 30 independent estimates of the sample standard deviation for both the section estimates and the jackknife estimates.

2. Likewise, the 300 estimates of the standard deviation were sectioned into 30 independent sections of 10 estimates of the standard deviation each. These 10 standard deviation estimates were averaged to get a single estimate of the standard deviation for each section. Thus there were 30 independent estimates of the standard deviation from the estimator, for both the section estimator and the jackknife estimator.

3. For each of the 30 sections, the mean of the 10 section or jackknife estimates of the standard deviation from step 2 was subtracted from the sample estimate of the standard deviation from step 1 to yield 30 independent estimates of the difference.

If the section or jackknife estimator is a reliable estimate of the sample standard deviation, then the difference of the sample standard deviation and the section or jackknife estimate of the standard deviation should be zero.

Note that while the same data is used for all of the section and jackknife estimators so that there is no independence between the different estimators, the 30 estimates of the difference for a single estimator i.e., the section estimate with $n = 25$ or the 2-fold jackknife are independent. Figure 5 has boxplots of the differences for both the section estimates (top graph) and the jackknife estimates (bottom graph).

The top graph in Figure 5, of the section estimator, shows that the sample mean for the smaller n is within one standard error of zero. When n is increased to 250 and 500, where the section estimates of the standard deviation are more variable because of the small m, the means of the differences, .00140 and .00300, are still within three standard errors of zero. This shows that section estimator of the standard deviation of the section point estimate is a reliable estimate of the sample standard deviation of the point estimate.

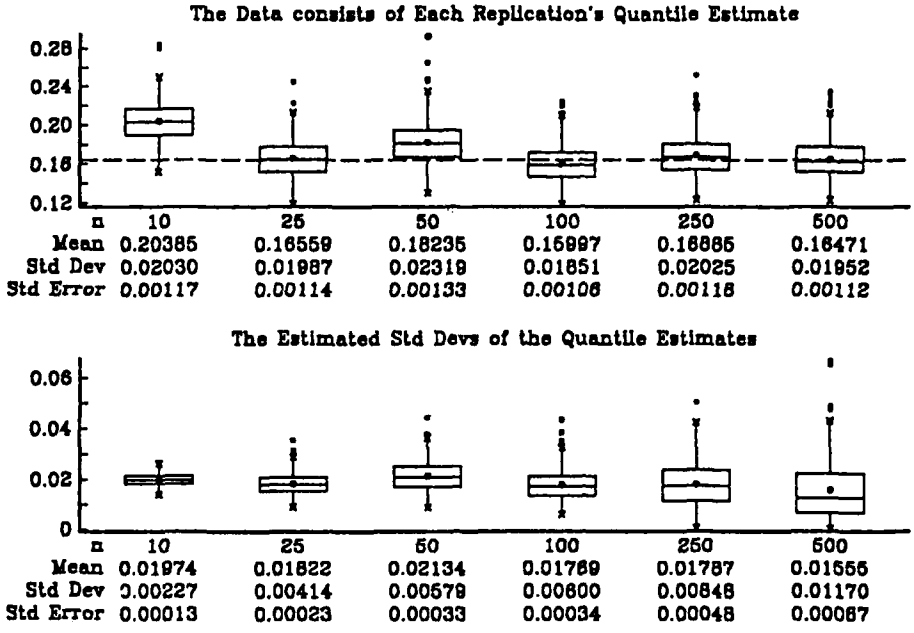

Figure 3: Boxplots of section point estimates of $y_{.95}$ (top) and section estimates of the standard deviation of the point estimates (bottom) for 300 replications of $N = 1000$ samples and varying n.

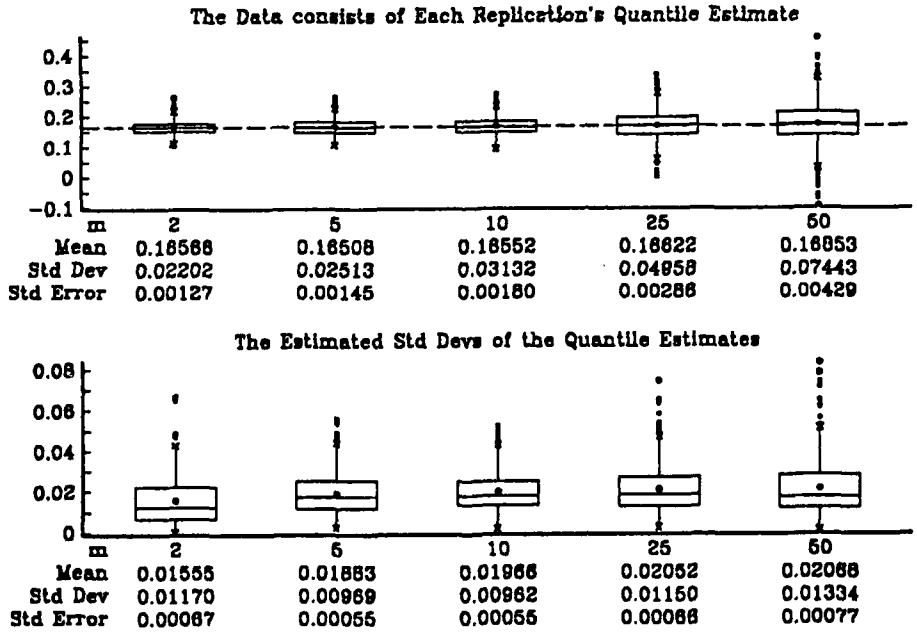

Figure 4: Boxplots of m-fold jackknife point estimates of $y_{.95}$ (top) and m-fold jackknife estimates of the standard deviation of the point estimates (bottom) for 300 replications of $N = 1000$ samples and varying m.

The bottom graph in Figure 5 shows the opposite for the jackknife estimator. For no m is the mean of the differences within three standard errors of zero. If one tests, for each m, the normality of the differences for the jackknife estimates, one can not reject at the .95 confidence level the hypothesis that the differences have a normal distribution. For each m, the .95 confidence interval for the mean of the fitted normal distribution does not include zero. Thus the jackknife estimate of the standard deviation of a jackknifed quantile estimate is a biased and unreliable estimate. We feel this is strong evidence for not using the jackknife technique for estimating quantiles and the variance of the quantile estimate.

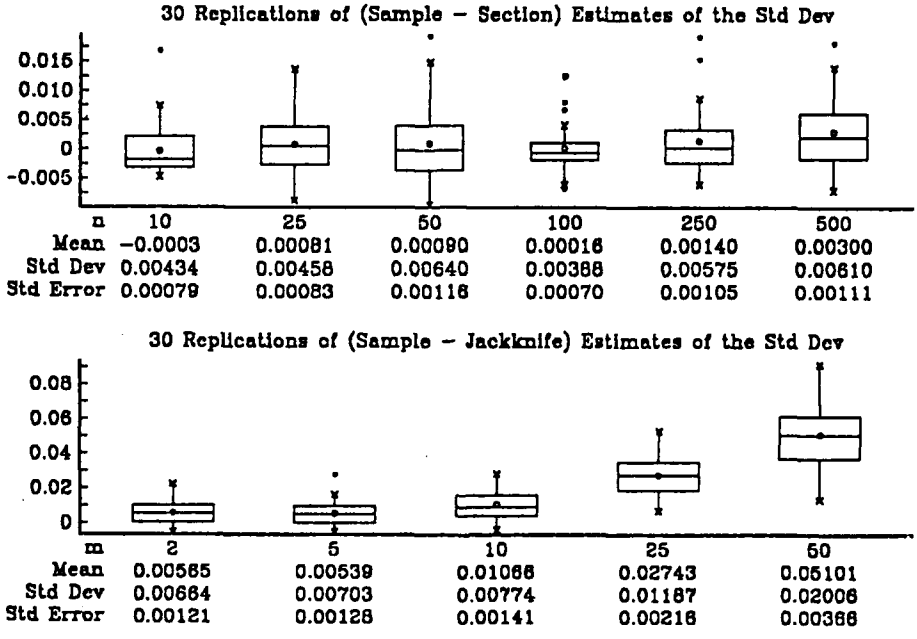

Figure 5: Boxplots of differences between estimates of the sample standard deviation of the point estimate and the section (top) and m-fold jackknife (bottom) estimates of the standard deviation of the point estimate based on 30 sections of $M = 300$ independent replications of $N = 1000$ samples each.

6.4 Comparing the Crude, Linearly Controlled and Nonlinearly Controlled Estimators

The crude, linearly controlled and nonlinearly controlled estimators will be compared both graphically and numerically. Now the number of replications is $M = 20$ and the number of samples in each replication is fixed at $N = 1000$. The section estimator will be used for all three estimators. For the nonlinearly controlled estimator, the monotone transformation will be the scaled power transformation so that the control function will be

$$\hat{y}'_\alpha(n) = \hat{y}_\alpha(n) - \theta_1 \left\{ \frac{\hat{x}_\alpha(n)^{\theta_2} - 1}{\theta_2} - \frac{x_\alpha^{\theta_2} - 1}{\theta_2} \right\}.$$

6.4.1 Comparison When the Sample Size $N = 1000$

Figure 6 shows the performance of the three estimators as triplets of boxplots for $n = 25$, 100, 250, and 500. *In each of the graphs that follow, the left boxplot of the triple is the crude estimate, the middle boxplot of the triple is the linearly controlled estimate and the right boxplot of the triple is the nonlinearly controlled estimate.* The statistics under each graph are the respective means of the data in the boxplot for the crude, linearly controlled and nonlinearly controlled estimators.

The boxplots in the top graph of Figure 6 contain the final quantile estimates for each of the estimators. This graph shows the effect of a control function that is biased because of the use of the asymptotic expected value. Without the biased control function each of the boxplots would look virtually the same because the control function would be mean zero and so would not change the expected value of the point estimate. The bias in the control function tends to reduce the bias of the point estimate with the exception of the linearly controlled estimate at $n = 25$.

The boxplots in the bottom graph of Figure 6 contain the section estimates of the standard deviation of the point estimators. One can see that as n increases, the mean of the estimated standard deviation of the linearly controlled estimate decreases, from .01123 to .00391, while the mean of the estimated standard deviation for the nonlinear control increases, once n is greater than 100, from .00241 to .00374, until the values for the linear control and the nonlinear control are about the same. In fact, the estimator that minimizes the variance can be seen to be the nonlinearly controlled estimator at $n = 100$ with a value of .00241. It is also clear that when n is large at 250 and 500, the small m of 4 and 2 causes higher variance in the estimates of the standard deviation.

The top graph in Figure 7 combines the two graphs from Figure 6, the bias and the variance, in that it contains the estimated mean square error of the estimators. It can be seen with this graph that the estimator that minimizes the mean square error is again the nonlinearly controlled estimator at $n = 100$ with a value of .00005. In fact the estimated mean square error for this estimator is under one-half of the best mean square error for the linear control of .00013 that is at $n = 250$. At $n = 500$ the values are the same, .00029, since there are only 2 quantile estimates with which to work. The other factor affecting the nonlinear control besides having only 2 quantile estimates to work with is that at $n = 500$ the joint distribution of the crude estimate and the control estimate is closer to multivariate normal than at $n = 100$.

The bottom graph in Figure 7 is a summary of the percent variance reduction achieved by the various estimators. The percent variance reduction for each estimator is computed using the estimate of the variance of the crude estimate which is why the value for the crude estimator is 0. This graph again highlights the effectiveness of the nonlinearly controlled estimator at smaller n. The highest percent variance reduction is .97568, which is actually achieved at $n = 25$ and not $n = 100$ because the percent variance reduction is a relative measure and the crude estimator at $n = 25$ had higher variance than the crude estimator at $n = 100$. This graph also points out the high variability of the variance reduction for large n as the number of quantile estimates becomes small.

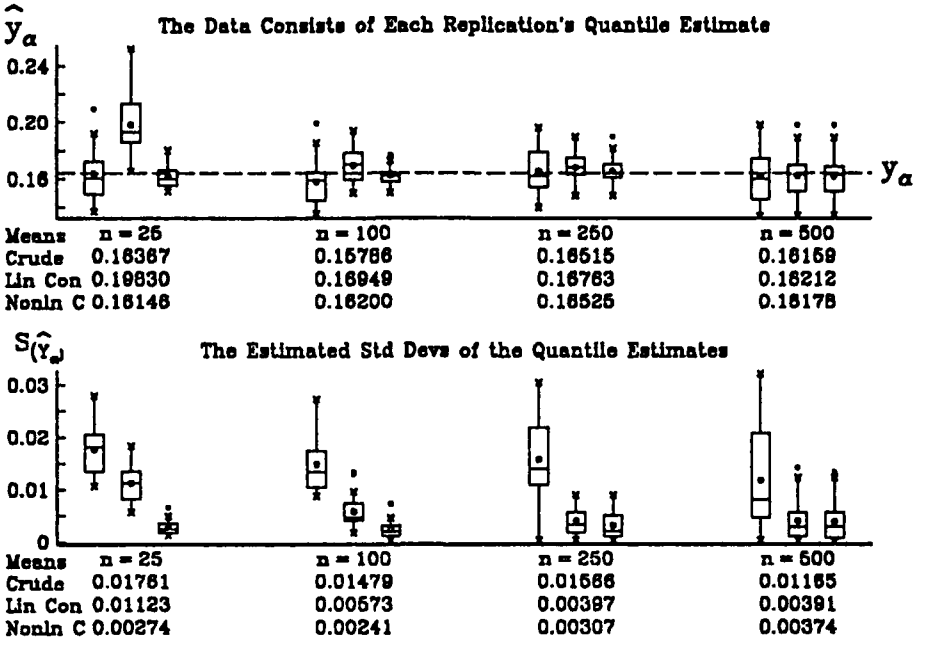

Figure 6: Boxplots of section crude, linearly controlled and nonlinear controlled estimators showing the point quantile estimates of $y_{.95}$ (top) and the estimates of the standard deviation of the point estimates (bottom) from $M = 20$ independent replications of $N = 1000$ for varying n.

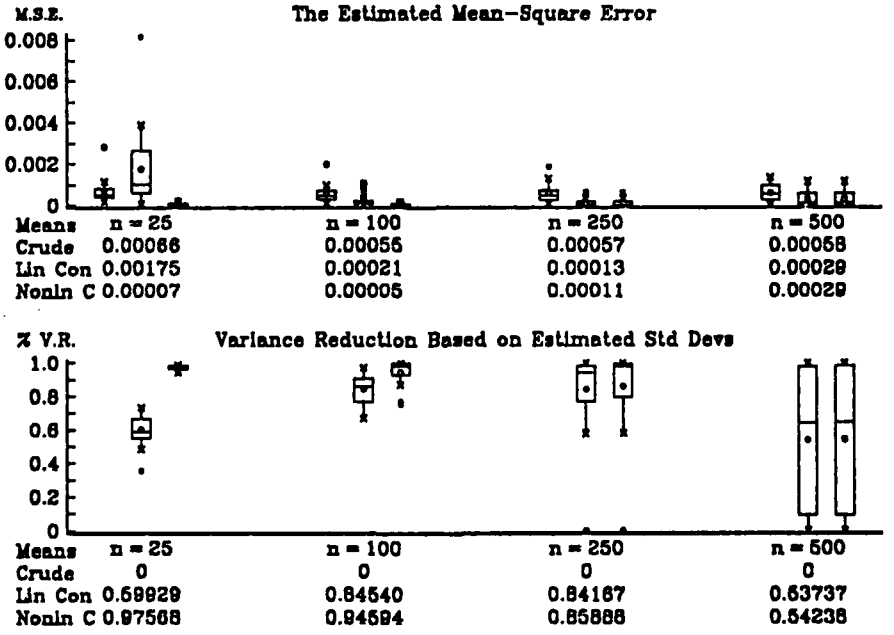

Figure 7: Boxplots of section crude, linearly controlled and nonlinear controlled estimators showing the estimated mean square error (top) and percent variance reduction (bottom) from $M = 20$ independent replications of $N = 1000$ for varying n.

6.4.2 Comparison When the Sample Size $N = 5000$

The next pairs of graphs, Figures 8 and 9 are identical in nature to the graphs for $N = 1000$ only now the data is from estimates made from a sample size of $N = 5000$. The number of samples used to compute each section estimate n is unchanged so increasing the sample size only increases m, the number of quantile estimates. The larger m greatly reduces the problem of high variability of the estimates caused by having only 2 quantile estimates with which to work at $n = 500$.

In the top graph of Figure 8, increasing m has slightly improved the bias of the mean of the nonlinearly controlled estimates so that it is now less than the bias of the crude estimate for each n. At the same time the bias of the mean of the linearly controlled estimates has increased. A more significant impact of increasing m, shown in the bottom graph, is the drop in the estimated standard deviations for all estimators as compared to $N = 1000$. The variability of the estimates of the standard deviation has decreased as well.

The mean square errors of the top graph in Figure 9 show again the nonlinear control at $n = 100$ does better than the best linearly controlled estimate. However, as n increases, one can lose the effectiveness of the nonlinear control as both the number of quantile estimates decreases and the quantile estimates approach multivariate normality. The impact of increasing N and m from Figure 7 is seen in the bottom graph of Figure 9 as the variability of the estimate of the percent variance reduction is greatly reduced.

7 SUMMARY

Nonlinear controls have been seen to be effective in improving the variance reduction over linearly controlled estimates of the mean. Sectioning is a useful procedure for computing point estimates for quantiles along with an estimate of the variance of the point estimate. The jackknife is not a useful procedure as the jackknife estimate of the variance of the jackknife point estimate is unreliable. Controlling quantiles with nonlinear controls is analytically tractable if the nonlinear transformations of the control quantile estimator are limited to strictly monotone functions. With this restriction, one can approximate the expected value of the transformed quantile estimator with its asymptotic expected value, namely the transformed value of the true quantile for the control. The approximation induces additional bias into the control function. However use of a biased control function can reduce the first order bias in the controlled estimate.

Finally, when one is considering the choice of m and n to use for the sectioning estimator, one must keep n small and avoid approaching the asymptotic multivariate normal distribution. As the joint distribution of the crude estimate of the quantile of interest and the control quantile estimate approaches multivariate normality, the effectiveness of the nonlinear control reduces to that of the linear control.

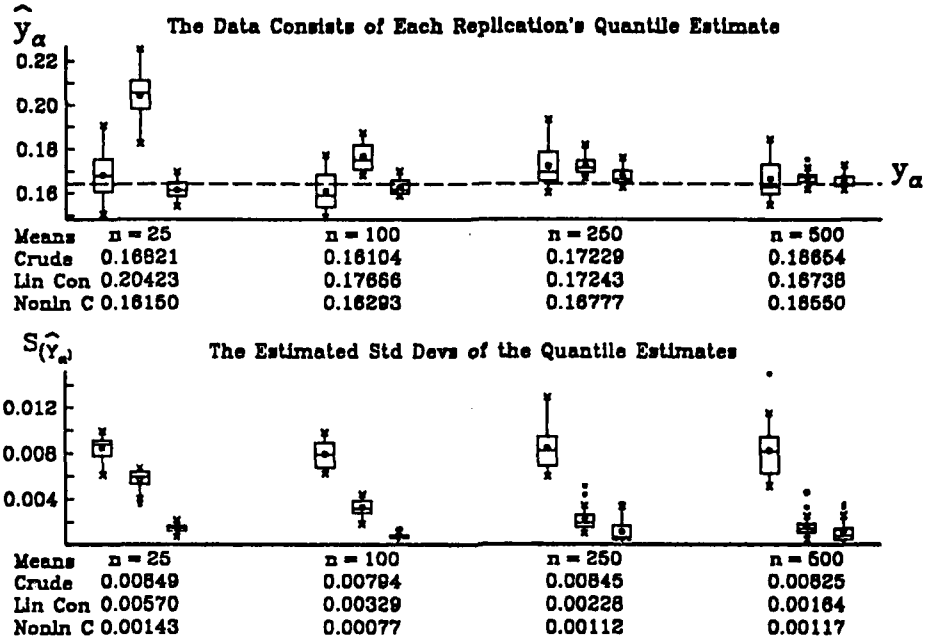

Figure 8: Boxplots of section crude, linearly controlled and nonlinear controlled estimators showing the point quantile estimates of $y_{.95}$ (top) and the estimates of the standard deviation of the point estimates (bottom) from $M = 20$ independent replications of $N = 5000$ for varying n.

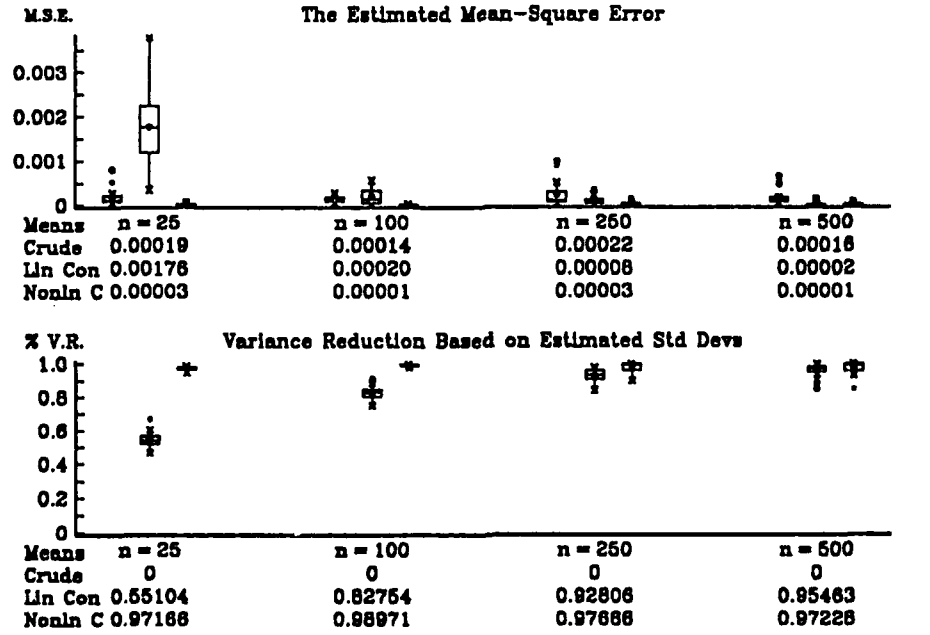

Figure 9: Boxplots of section crude, linearly controlled and nonlinear controlled estimators showing the estimated mean square error (top) and percent variance reduction (bottom) from $M = 20$ independent replications of $N = 5000$ for varying n.

ACKNOWLEDGMENTS

The research of P.A.W. Lewis was supported at the Naval Postgraduate School under an Office of Naval Research Grant.

We are indebted to Dr. P.D. Welch for permission to use the statistical computing package GRAFSTAT under a test-site agreement with IBM Research.

BIBLIOGRAPHY

Beale, E., "Regression: A Bridge Between Analysis and Simulation", *The Statistician*, v. 34, pp. 141–154, 1985.

Breiman, L. and Friedman, J. H., "Estimating Optimal Transformations for Multiple Regression and Correlation", *Journal of the American Statistical Association*, v. 80, no. 391, pp. 580–619, September 1985.

Chambers, J. M., Cleveland, W. S., Kleiner, B., and Tukey, P. A., *Graphical Methods for Data Analysis*, Duxbury Press, 1983.

David, H. A., *Order Statistics*, John Wiley & Sons, Inc, 1970.

Efron, B. and Gong, G., "A Leisurely Look at the Bootstrap, the Jackknife and Cross-validation", *The American Statistician*, v. 37, no. 1, pp. 36–48, February 1983.

Gallant, A. R., *Nonlinear Regression*, John Wiley & Sons, 1987.

Glynn, P. W. and Whitt, W., "Indirect Estimation via $L = \lambda W$", *Operations Research*, v. 37, no. 1, pp. 82–103, 1989.

Heidelberger, P. and Lewis, P. A. W., "Regression-adjusted Estimates for Regenerative Simulations, with Graphics", *Communications of the ACM*, v. 24, no. 4, pp. 260–273, April 1981.

Hsu, J. C. and Nelson, B. L., "Control Variates for Quantile Estimation", In *Proceedings of the 1987 Winter Simulation Conference*, Thesen, A., Grant, H., and Kelton, W. D., editors, pp. 342–346, 1987.

Kendall, M. and Stuart, A., *The Advanced Theory of Statistics*, 4th ed., v. 1, Macmillan, 1977.

Lancaster, H. O., "Kolmogorov's Remark on the Hotelling Canonical Correlation", *Biometrika*, v. 53, no. 3 and 4, pp. 585–590, 1966.

Lavenberg, S. S., Moeller, T. L., and Welch, P. D., "Statistical Results on Control Variables with Application to Queueing Network Simulation", *Operations Research*, v. 30, no. 1, pp. 182–202, January and February 1982.

Lewis, P. A. W. and Orav, J., *Simulation Methodology for Statisticians, Operations Analysts, and Engineers*, v. 1, Wadsworth & Brooks/Cole, 1989.

Lewis, P. A. W., Ressler, R. L., and Wood, R. K., "Variance Reduction using Nonlinear Controls and Transformations", *Communications in Statistics, Part B*, v. 18, no. 2, pp. 655–672, 1989.

Miller, R. G., "The Jackknife - A Review", *Biometrika*, v. 61, no. 1, pp. 1–15, 1974.

Mood, A. M., Graybill, F. A., and Boes, D. C., *Introduction to the Theory of Statistics*, 3rd ed., McGraw Hill, 1974.

Nelson, B. L., *Control-Variate Remedies*, Department of Industrial and Systems Engineering, The Ohio State University, Report 1988-004, October 1988.

Rubinstein, R. Y. and Marcus, R., "Efficiency of Multivariate Control Variates in Monte Carlo Simulation", *Operations Research*, v. 33, no. 3, pp. 661–667, May-June 1985.

Seila, A. F., "A Batching Approach to Quantile Estimation in Regenerative Simulation", *Management Science*, v. 28, pp. 573–581, 1982.

Serfling, R. J., *Approximation Theorems of Mathematical Statistics*, Wiley, New York, 1980.

Tocher, K. D., *The Art of Simulation*, The English Universities Press, Ltd, 1963.

Weiss, L., "On the Asymptotic Joint Normality of Quantiles from a Multivariate Distribution", *Journal of Research of the National Bureau of Standards*, v. 68B, no. 2, pp. 65–66, April-June 1964.

INITIAL DISTRIBUTION LIST

1. Library (Code 0142) 2
   Naval Postgraduate School
   Monterey, CA 93943-5000

3. Defense Technical Information Center 2
   Cameron Station
   Alexandria, VA 22314

4. Office of Research Administration 1
   Code 012A
   Naval Postgraduate School
   Monterey, CA 93943-5000

5. Prof. Peter Purdue, Code OR/Pd 1
   Naval Postgraduate School
   Monterey, CA 93943-5000

6. Prof. Peter A. W. Lewis 105
   Code OR/Lw
   Naval Postgraduate School
   Monterey, CA 93943-5000