Regression estimators in simulation

Hopmans, A.C.M.; Kleijnen, J.P.C.

Publication date:
1977

REGRESSION ESTIMATORS IN SIMULATION

by

ANTON C. M. HOPMANS

and

JACK P. C. KLEIJNEN

Research memorandum

TILBURG UNIVERSITY
DEPARTMENT OF ECONOMICS
Postbus 90135 - 5000 LE Tilburg
Netherlands

REGRESSION ESTIMATORS IN SIMULATION

by

ANTON C.M. HOPMANS
(Dr. Neher Laboratorium, PTT, Leidschendam)

and

JACK P.C. KLEIJNEN
(Department of Economics & Business, Katholieke Hogeschool Tilburg)

December 1977

Contents

	page
Abstract	1
1. Introduction	1
2. Some preliminaries	5
3. Possible bias of $\bar{B}$ and $\hat{B}_\alpha$	8
3.1. The crude estimator $\bar{B}$	10
3.2. The regression estimator $\hat{B}_\alpha$	12
4. Linear approximation to f(A)	12
4.1. The crude estimator $\bar{B}$	13
4.2. The regression estimator $\hat{B}_\alpha$	14
5. The variance of $\bar{B}$ and $\hat{B}_\alpha$	14
5.1. The crude estimator $\bar{B}$	15
5.1. The regression estimator $\hat{B}_\alpha$	18
6. Conclusion and application	19
Notes	22
References	24

Keywords and phrases: simulation, Monte Carlo, regression, least squares, control variates, variance reduction, queuing systems, networks.

CR Categories 3.24, 4.6, 5.5, 8.1, 8.3

REGRESSION ESTIMATORS IN SIMULATION

by

ANTON C.M. HOPMANS
(Dr. Neher Laboratorium, PTT, Leidschendam, Neth.)

and

JACK P.C. KLEIJNEN
(Katholieke Hogeschool, Tilburg, Netherlands)

Abstract

Dividing a simulation run into subruns yields average input values per subrun which deviate from their known expectation. This information can be used to improve the estimated simulation response: control variates or regression sampling. To derive the statistical properties of the new estimator, regression analysis is examined for stochastic independent variables and misspecified regression models. It is shown that the usual, crude estimator is biased.
Assuming a local linear approximation, the crude estimator remains biased "ex-post", whereas the regression estimator becomes unbiased. Moreover, the variance of the regression estimator is smaller under each of three intuitivily acceptable conditions.

1. INTRODUCTION

In this section we shall briefly describe the "background" problem that lead to our interest in control variates, and define control variates and its equivalent, regression sampling. Control variates is a procedure that was applied "spontaneously" by one of the authors

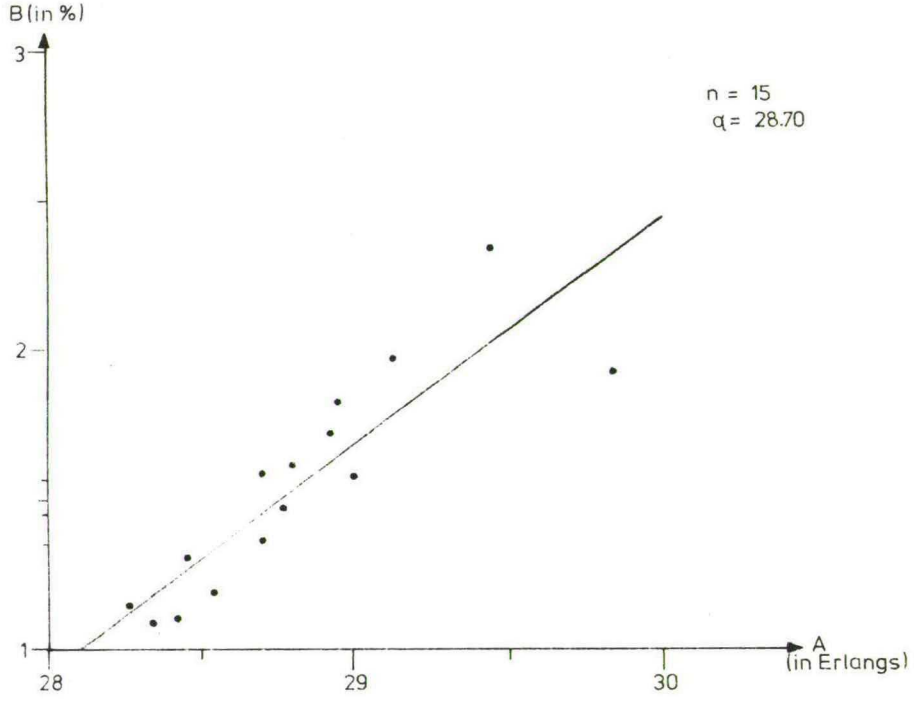

FIG. 1. Estimated blocking B versus arrival intensity A

in a simulation experiment with a service network (called a "grading") as experienced in PTT telephone-exchange systems; see Bear (1976) for a description of such systems. The purpose of the simulation was to estimate the steady-state probability of blocking, i.e., a telephone call (customer) finds all lines (servers) occupied. To estimate the accuracy (reliability) of the simulated blocking probability, the standard deviation σ (standard error) of the simulation response is needed. One approach for estimating this σ is to cut the total simulation run into subruns of predetermined length. (Assuming independence or testing for independence of subrun responses, straightforward calculation of the estimate $\hat{\sigma}$ is then possible; see Kleijnen (1977) for details on this procedure and alternative procedures.) Control variates arise as follows.

Per subrun the average arrival rate of customers[1)] can be computed, denoted by $\underline{A}$ (stochastic variables are underlined to distinguish them from the values they can assume). The value A in a subrun will not be exactly equal to the theoretical, expected value, say α. This expected value is known as the simulation program samples from a known distribution of interarrival times (known input). Note that we are interested in estimating the blocking probability for $A = \alpha$ (desired input value), not for $A = \bar{A}$ (accidentally sampled value). If in subrun i ($i = 1,\ldots n$), relatively many customers arrive ($A_i > \alpha$), then we expect this subrun to have a relatively high estimated blocking probability, say B_i. Hence, we may regress blocking probabilities B_i on traffic intensities A_i, as in FIG. 1. In order to estimate the blocking percentages at the value we are interested in ($A = \alpha$), we can use the regression model

$$\underline{B}_i = \beta_0 + \beta_1 \cdot \underline{A}_i + \underline{u}_i \qquad (1.1)$$

Applying the least squares procedure to estimate β_0 and β_1 yields

$$\hat{\underline{\beta}}_\alpha = \hat{\underline{\beta}}_0 + \hat{\underline{\beta}}_1 \cdot \alpha \qquad (1.2)$$

Note that the regression model (1.1) is extremely simple, except for

the fact that the independent variable itself is stochastic (since the simulation program yields stochastic traffic intensities per subrun).

The regression sampling estimator $\hat{\underline{B}}_\alpha$ of (1.2) can also be represented as a so-called control variate estimator; as we shall prove next. Define the control variate estimator $\tilde{\underline{B}}$ as

$$\tilde{\underline{B}} = \bar{\underline{B}} + \beta_c \, (\alpha - \bar{\underline{A}}) \tag{1.3}$$

with the standard definitions

$$\bar{\underline{B}} = \sum_1^n \underline{B}_i / n \ , \quad \bar{\underline{A}} = \Sigma \underline{A}_i / n \tag{1.4}$$

and β_c being a correction coefficient to which we shall return. $\bar{\underline{B}}$ is the standard (crude) response of the simulation run when not using regression or control variates (overall mean of all subruns). If, for instance, $\alpha > \bar{A}$ (too low $\bar{A}$), then intuitively we wish to correct $\bar{\underline{B}}$ such that $\tilde{B} > \bar{B}$. This is realized by selecting a positive value for the correction coefficient β_c. It is simple to derive, see Kleijnen (1975, p. 140), that the variance of $\tilde{\underline{B}}$ is minimized by selecting the optimal value

$$\beta_c^* = \rho \, \sigma_B / \sigma_A \tag{1.5}$$

where σ_A^2 and σ_B^2 are the variances of $\underline{A}$ and $\underline{B}$, and ρ is the correlation coefficient between $\underline{A}$ and $\underline{B}$.

However, the optimal correction coefficient in control variates β_c^* is identical to the least squares coefficient β_1 in the regression model (1.1), as can be verified in any textbook on regression analysis. Hence, we use as an estimator[2)]

$$\hat{\underline{\beta}}_c^* = \hat{\underline{\beta}}_1 \tag{1.6}$$

But then it is simple to derive that both estimators, optimal control variates and regression sampling, are identical: From the least squares procedure it follows that the regression line passes through the "center of gravity"; see Johnston (1972, p. 16). In other words, if $A = \bar{A}$, then

$\hat{B} = \bar{B}$, or

$$\underline{\hat{B}}_{\bar{A}} = \underline{\hat{\beta}}_0 + \underline{\hat{\beta}}_1 . \underline{\bar{A}} = \underline{\bar{B}} \tag{1.7}$$

This implies that the **intercept** $\hat{\beta}_0$ satisfies

$$\underline{\hat{\beta}}_0 = \underline{\bar{B}} - \underline{\hat{\beta}}_1 . \underline{\bar{A}} \tag{1.8}$$

Substitution into (1.2) yields

$$\underline{\hat{B}}_\alpha = (\underline{\bar{B}} - \underline{\hat{\beta}}_1 . \underline{\bar{A}}) + \underline{\hat{\beta}}_1 . \alpha = \underline{\bar{B}} + \underline{\hat{\beta}}_1 (\alpha - \underline{\bar{A}}) = \underline{\tilde{B}}^* \tag{1.9}$$

where the last equality follows from (1.3) and (1.6). The above derivations are well-known in the literature on variance reduction techniques in simulation, and in other forms of sampling experiments such as sample surveys. However, the statistical properties of the control variate estimator are not well understood, even though the technique has been applied in a number of simulation studies; see Kleijnen (1975, pp. 138-164). In the remainder we shall examine the possible bias and the variance of control variates. We shall also discuss an aspect much neglected in the literature, namely how to construct confidence intervals for the new estimator.

2. Some preliminaries

We mentioned that the regression model (1.1) is extremely simple except for the stochastic character of the independent variable. It can be found in, e.g., Johnston (1972, p. 40) that

$$\text{var}\,(\underline{\hat{B}}_\alpha) \equiv \text{var}\,(\underline{\hat{B}} | \underline{A} = \alpha) = \sigma_v^2 \{\frac{1}{n} + \frac{(\alpha - \bar{A})^2}{\Sigma (A_i - \bar{A})^2}\} \tag{2.1}$$

Note that the variance of $\underline{\hat{B}}$ increases as we move away from the center of gravity. What are the assumptions of (2.1)?

First, consider the general model

$$\vec{\underline{y}} = \vec{X}\vec{\beta} + \vec{\underline{u}} \tag{2.2}$$

where matrices and vectors are denoted by $\rightarrow$. Its least squares estimators are

$$\vec{\hat{\underline{\beta}}} = (\vec{X}'\vec{X})^{-1}\,\vec{X}'\vec{\underline{y}} = \vec{W}\,\vec{\underline{y}} \tag{2.3}$$

where we introduced the shortland notation $\vec{W} = (\vec{X}'\vec{X})^{-1}\,\vec{X}'$. In Scheffé (1964, p. 8) the following general result can be found. If we have a stochastic vector $\vec{\underline{y}}_1$ with covariance matrix Ω_1 then the linear transformation $\vec{\underline{y}}_2 = \vec{A}\,\vec{y}_1$ has covariance matrix $\vec{\Omega}_2 = \vec{A}\,\vec{\Omega}_1\,\vec{A}'$. Hence the variance-covariance matrix $\vec{\Omega}$ of $\vec{\hat{\underline{\beta}}}$ is

$$\vec{\Omega}_\beta = \vec{W}\,\vec{\Omega}_y\,\vec{W}' \tag{2.4}$$

The predicted value of $\underline{y}$ for, say, the row vector of independent variables $\vec{x}' = \vec{x}_0'$, is

$$\hat{\underline{y}} = \vec{x}_0'\,\vec{\hat{\underline{\beta}}} \tag{2.5}$$

Hence, similar to (2.4), we have

$$\text{var}\,(\hat{\underline{y}}\,|\,\vec{x} = \vec{x}_0) = \vec{x}_0'\,\vec{\Omega}_\beta\,\vec{x}_0 \tag{2.6}$$

Returning to (2.1) we conclude that this variance expression does not depend on the following assumptions[3]:

(i) Normality of the observations $\vec{\underline{y}}$.[4]

(ii) Correctness of the fitted (linear) regression model. Obviously the bias of $\hat{\underline{y}}$ does depend on this assumption[5].

Eq. (2.1) is based on the next assumptions.

(iii) Non-stochastic independent variables, since (2.4) follows from (2.3) assuming a deterministic $\vec{W}$.

(iv) Constant variances $\sigma_i^2 = \sigma_v^2$ where

$$\text{var}\,(\underline{B}\,|\,A = A_i) = \sigma_i^2 = \sigma_v^2 \qquad (i = 1,\ldots,n) \tag{2.7}$$

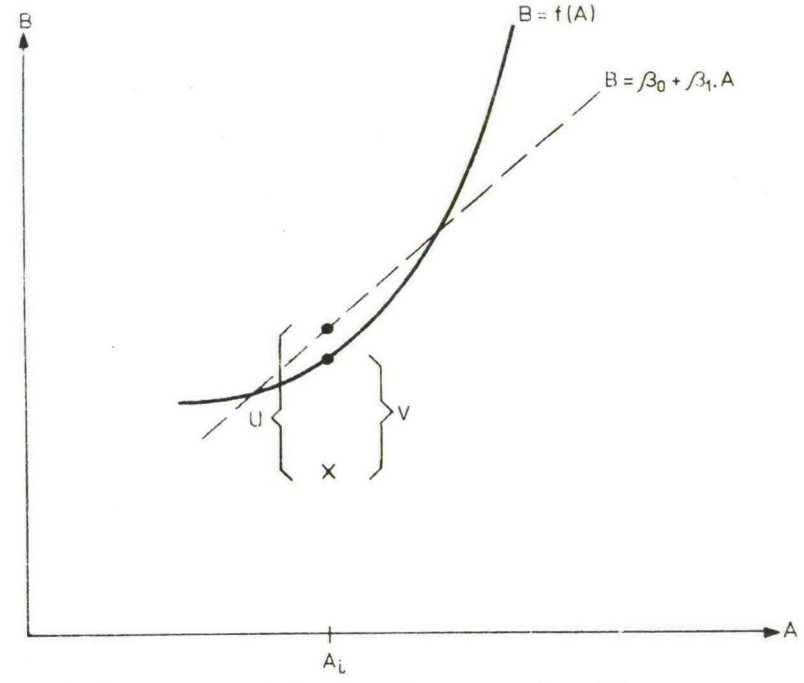

FIG. 2. Noise around true and assumed model

where σ_i^2 is the variance of $\underline{B}$ around the (unknown) true model, say,

$$\underline{B} = f(\underline{A}) \tag{2.8}$$

So, we have to distinguish between $\underline{v}$, the noise of the true model, and $\underline{u}$, the noise of the assumed (linear) model (1.1), which noise may be biased, i.e. $\mathcal{E}(\underline{u}) \neq 0$. See also FIG. 2. Iff the model is correctly specified, then we can estimate the variance σ_u^2 by the mean squared residuals $\underline{e}(\equiv \underline{B}-\hat{\underline{B}})$:

$$\hat{\sigma}_e^2 = \sum_1^n (\underline{B}_i-\hat{\underline{B}}_i)^2/(n-2) \tag{2.9}$$

Next we turn to our original problem where the independent variable is stochastic and the regression model may be misspecified.

3. Possible bias of $\bar{\underline{B}}$ and $\hat{\underline{B}}_\alpha$

If subrun i has one specific value A_i for its traffic intensity, then the blocking percentage $\underline{B}_i$ can still assume various values (depending on the order in which "customers" arrive, and other stochastic factors). Each subrun is subject to the same probability law so that we can write $\underline{A}_i = \underline{A}$ and $\underline{B}_i = \underline{B}$. As FIG. 2 demonstrated, we have

$$\mathcal{E}(\underline{B}|\underline{A} = A) = f(A) \tag{3.1}$$

By definition the unconditional expectation satisfies

$$\mathcal{E}(\underline{B}) = \mathcal{E}_A\{\mathcal{E}(\underline{B}|\underline{A} = A)\} = \mathcal{E}_A\{f(\underline{A})\} \tag{3.2}$$

Unless f(A) is linear, we know that

$$\mathcal{E}\{f(\underline{A})\} \neq f\{\mathcal{E}(\underline{A})\} = f(\alpha) \tag{3.3}$$

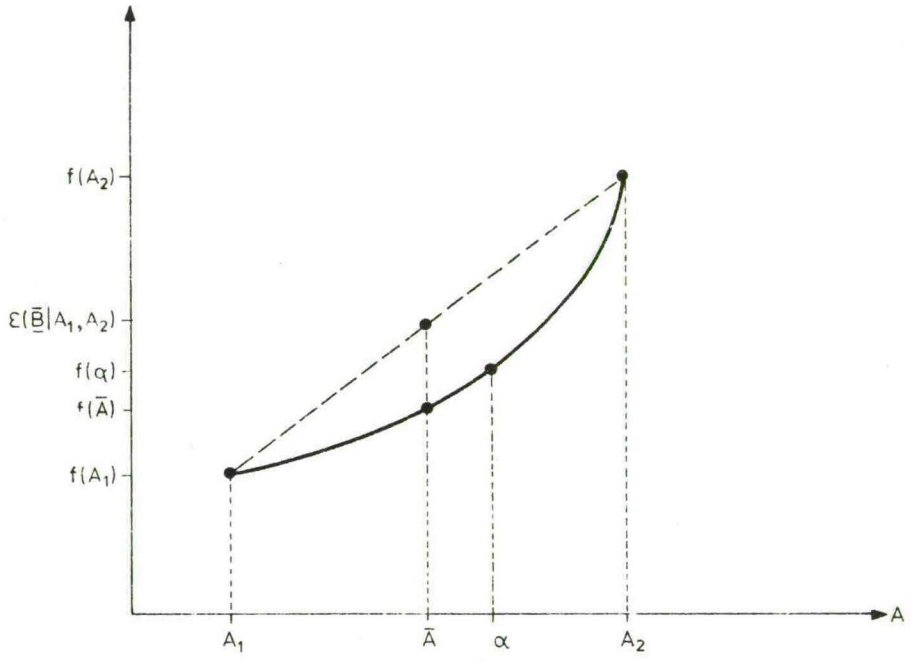

FIG. 3. Bias of crude estimator $\underline{\bar{B}}$

3.1. The crude estimator $\bar{B}$

Consider the first estimator $\bar{B}$, i.e., the "crude" estimator defined in (1.4). Obviously

$$\&(\bar{B}|A_1,\ldots,A_n) = \&(\Sigma B_i/n|A_1,\ldots,A_n) = \frac{1}{n}\Sigma\&(B_i|A_i) \qquad (3.4)$$

where the last equality holds since the B_i are supposed to be independent after dividing the total simulation run into n subruns of appropriate length; see section 1. In other words, B_i does not depend on A_j $(j \neq i)$. Substituting (3.1) into (3.4), and defining $\vec{A} = (A_1,\ldots,A_n)$, yields

$$\&(\bar{B}|A) = \frac{1}{n}\Sigma f(A_i) \neq f\{\Sigma(A_i/n)\} = f(\bar{A}) \qquad (3.5)$$

We also refer to the illustration in FIG. 3 where n = 2. From (3.5) and (3.2) we derive

$$\begin{aligned}\&(\bar{B}) &= \underset{A}{\&}\{\frac{1}{n}\sum_1^n f(A_i)\} = \frac{1}{n}\sum_1^n \&\{f(A_i)\} \\ &= \&\{f(A)\} \\ &\neq f\{\&(A)\} = f(\alpha) \end{aligned} \qquad (3.6)$$

In other words, even the simple estimator $\bar{B}$ is a biased estimator, since the traffic intensity per subrun A_i is stochastic instead of being controlled at the level α. See also FIG. 3 (where we use as an illustration a situation where $\alpha > \bar{A}$). This figure illustrates the intuitive notions that the bias reduces for

(i) increased subrun-length so that the subrun traffic intensities A_i tend to be closer to $\alpha = \&(A)$,

(ii) more subruns so that the probability of "many" A-values far from α, decreases. Then we can better approximate f(A) locally by a linear function.

Given the simulation esperiment, the (ex-post) bias follows from (3.5) and is

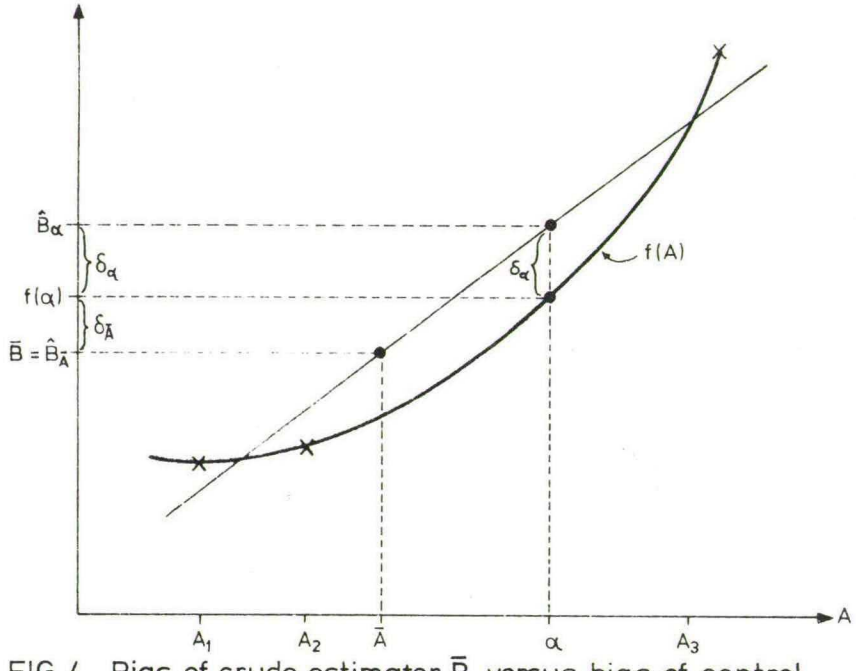

FIG. 4. Bias of crude estimator $\bar{B}$ versus bias of control variate $\hat{B}_\alpha$

$$\text{Bias } (\underline{\bar{B}}|\vec{A}) = \frac{1}{n}\Sigma f(A_i) - f(\alpha) \tag{3.7}$$

Before the simulation run is executed, the (ex-ante) bias is

$$\text{Bias } (\underline{\bar{B}}) = \mathcal{E}\{f(\underline{A})\} - f(\alpha) \tag{3.8}$$

3.2. The regression estimator $\underline{\hat{B}}_\alpha$

Next we consider the regression or optimal controle variate estimator $\underline{\hat{B}}_\alpha$ defined in (1.9). Its relation to the crude estimator $\underline{\bar{B}}$ is illustrated in FIG. 4, where we eliminated the noise terms $\underline{v}$ and took $n > 2$ so that the linear approximation does not give exact fit. In general, we have

$$\begin{aligned}\text{Bias } (\underline{\hat{B}}_\alpha|\vec{A}) &= \mathcal{E}(\underline{\hat{\beta}}_0 + \underline{\hat{\beta}}_1.\alpha|\vec{A}) + \\ &\quad - (\beta_0^{\vec{A}} + \beta_1^{\vec{A}}\alpha + \delta_\alpha^{\vec{A}}) = -\delta_\alpha^{\vec{A}}\end{aligned} \tag{3.9}$$

where $\delta_\alpha^{\vec{A}}$ denotes the deviation between the true value $f(\alpha)$ and the linear approximation based on the observed $\vec{A} = (A_1,\ldots,A_n)$, evaluated at the point α. In (3.9) we use the definitions

$$\beta_0^{\vec{A}} = \mathcal{E}(\underline{\hat{\beta}}_0|\vec{A}), \qquad \beta_1^{\vec{A}} = \mathcal{E}(\hat{\beta}_1|\vec{A}) \tag{3.10}$$

Note that given $\vec{A}$, many B_i-values remain possible (see eq. 2.7), which may yield many values for $\hat{\beta}_0$ and $\hat{\beta}_1$. FIG 4 also shows the crude estimator $\underline{\bar{B}}$, which equals $\underline{\hat{B}}$ evaluated at $\bar{A}$. Hence, this estimator has bias because $f(A)$ is approximated linearly, and because we measure at $\bar{A}$ instead of α (see also FIG. 3). Since the formulas of this section do not provide further insight, we shall use a linear approximation to $f(A)$ in the next section.

4. Linear approximation to f(A)

In this section we assume that $\mathcal{E}(\underline{B}|A) = f(A)$ can be approximated by a linear function "locally", i.e., in the neighbourhood of the point of

interest α. Hence, we assume

$$\mathcal{E}(\underline{B}|A) = f(A) = \beta_0 + \beta_1 . A \qquad (4.1)$$

provided A is the neighbourhood of α, say

$$\alpha^- < A < \alpha^+ \qquad (4.2)$$

Note that such a linear approximation is not a drastic simplification when compared to the usual situation in experimental design. In such designs a factor like A is varied over a wide range in order to detect whether A has any effect on the response. In our case the expected value for each $\underline{A}_i$ remains α, and the longer the subruns are, the closer the A_i are to α. In the remainder of this paper we assume that only A-values satisfying (4.2) are used; practically speaking we assume that the experimenter rejects "obvious" outlying observations on $\underline{A}$.
The formulas derived for a general function f(A) in the preceding section, are then replaced by the following results.

4.1. The crude estimator $\bar{\underline{B}}$

Eq. (3.3) is replaced by

$$\mathcal{E}\{f(\underline{A})\} = \mathcal{E}(\beta_0 + \beta_1 \underline{A}) = \beta_0 + \beta_1 \alpha \qquad (4.3)$$

Eq. (3.5) becomes

$$\mathcal{E}(\bar{\underline{B}}|\underline{A}) = \frac{1}{n} \Sigma(\beta_0 + \beta_1 A_i) = \beta_0 + \beta_1 \bar{A} \qquad (4.4)$$

Eq. (3.6) is replaced by

$$\mathcal{E}(\bar{\underline{B}}) = \mathcal{E}\{f(A)\} = \mathcal{E}\{\beta_0 + \beta_1 \underline{A}\} = \beta_0 + \beta_1 \alpha \qquad (4.5)$$

Hence (3.7) becomes

$$\text{Bias } (\underline{\bar{B}}|\vec{A}) = f(\bar{A}) - f(\alpha) = \beta_1.(\bar{A}-\alpha) \qquad (4.6)$$

so that in general ex-post bias will remain since $\bar{A} \neq \alpha$. However, the ex-ante bias is zero since eq. (3.8) becomes

$$\text{Bias } (\underline{\bar{B}}) = \mathcal{E}\{f(\underline{A})\} - f(\alpha) = \mathcal{E}(\beta_0 + \beta_1 A) +$$

$$- (\beta_0 + \beta_1\alpha) = \beta_1 \mathcal{E}(\underline{A}) - \beta_1.\alpha = 0 \qquad (4.7)$$

4.2. The regression estimator $\underline{\hat{B}}_\alpha$

In order to see how the estimated regression coefficients $\underline{\hat{\beta}}_0$ and $\underline{\hat{\beta}}_1$ depend on the independent variables $\vec{\underline{A}}$, we return to the general case, especially eq. (2.3). This yields

$$\mathcal{E}(\vec{\underline{\hat{\beta}}}|\vec{X}) = \mathcal{E}(\vec{W}\,\vec{y}|\vec{X}) = \vec{W}\,\mathcal{E}(\underline{\vec{y}}|\vec{X}) = (\vec{X}'\vec{X})^{-1}\vec{X}'.\vec{X}\vec{\beta} = \vec{\beta} \qquad (4.8)$$

where $\vec{W}$ is not stochastic as $\vec{X}$ is given. So if the assumed model is correct then any realization of the independent variables yields unbiased estimators ($\vec{X}$ is assumed to be non-singular). In our case this means

$$\mathcal{E}(\underline{\hat{\beta}}_0|\vec{A}) = \beta_0 \,, \qquad \mathcal{E}(\underline{\hat{\beta}}_1|\vec{A}) = \beta_1 \qquad (4.9)$$

Consequently

$$\mathcal{E}(\underline{\hat{B}}_\alpha|\vec{A}) = \mathcal{E}(\underline{\hat{\beta}}_0 + \underline{\hat{\beta}}_1\,\alpha|\vec{A}) = \beta_0 + \beta_1.\alpha \qquad (4.10)$$

Summarizing section 4, the regression estimator $\underline{\hat{B}}_\alpha$ has no bias whereas the crude estimator $\underline{\bar{B}}$ does show ex-post bias $\beta_1(\bar{A}-\alpha)$, assuming f(A) can be approximated linearly.

5. The variance of $\bar{B}$ and $\hat{B}_\alpha$

Having examined the bias in the preceding two sections, we next proceed to the variance of the two competing estimators. We shall consider two cases in this section, namely the general case f(A) and the special case of its linear approximation. We shall use a basic formula that can be found in, e.g., Keeping (1962, pp. 398-399):

$$\text{var}\ (\underline{x}) = \underset{y}{\&}\{\text{var}\ (\underline{x}|y)\} + \underset{y}{\text{var}}\{\&(\underline{x}|y)\} \tag{5.1}$$

5.1. The crude estimator $\underline{\bar{B}}$

Remember that the $\underline{B}_i$ were supposed to be independent because of the way subruns were formed. We further assumed that the $\underline{B}_i$ have constant variance σ_v^2; see eq. (2.7). Hence

$$\begin{aligned}\text{var}\ (\underline{\bar{B}}|\vec{A}) &= \text{var}\ (\tfrac{1}{n}\Sigma\underline{B}_i|\vec{A}) = \frac{1}{n^2}\sum_1^n \text{var}\ (\underline{B}_i|A_i) \\ &= \frac{1}{n^2}\,\Sigma\ \text{var}\{f(A_i) + \underline{v}_i\} = \frac{1}{n^2}\,\Sigma\ \text{var}\ (\underline{v}_i) = \sigma_v^2/n\end{aligned} \tag{5.2}$$

In other words, since the A_i-values are fixed, $\underline{B}_i$ and hence $\underline{\bar{B}}$ can vary only because of the noise $\underline{v}$.

Applying eq. (5.1) yields

$$\begin{aligned}\text{var}\ (\underline{\bar{B}}) &= \underset{\vec{A}}{\&}\{\text{var}\ (\underline{\bar{B}}|\vec{A})\} + \underset{\vec{A}}{\text{var}}\{\&(\underline{\bar{B}}|\vec{A})\} \\ &= \underset{\vec{A}}{\&}\{\sigma_v^2/n\} + \underset{\vec{A}}{\text{var}}\{\&(\tfrac{1}{n}\Sigma\underline{B}_i|\vec{A})\} \\ &= \sigma_v^2/n + \underset{\vec{A}}{\text{var}}\{\tfrac{1}{n}\Sigma\&(\underline{B}_i|A_i)\} \\ &= \sigma_v^2/n + \frac{1}{n^2}\,\underset{\vec{A}}{\text{var}}\{\Sigma\ f(A)\} \\ &= \sigma_v^2/n + \text{var}\{f(\underline{A})\}/n\end{aligned} \tag{5.3}$$

where the last two equalities hold since the $\underline{A}_i$ are identically and independently distributed. The first component of (5.3) corresponds with the variance around f(A) given A, and the second component corresponds with the variation along the curve f(A).

If we assume a linear approximation to f(A), then (5.3) becomes

$$\text{var } (\underline{\bar{B}}) = \sigma_v^2/n + \frac{1}{n} \text{ var } \{\beta_0 + \beta_1 A\}$$

$$= \sigma_v^2/n + \beta_1^2 . \text{var } (\underline{A})/n \qquad (5.4)$$

Observe that (5.4) could have been derived as follows. If we do not condition on the A_i-values, then the $\underline{B}_i$ are identically distributed with variance σ_B^2. Hence

$$\text{var } (\underline{\bar{B}}) = \text{var } (\Sigma \underline{B}_i/n) = \sigma_B^2/n \qquad (5.5)$$

If we use the linear model

$$\underline{B}_A = \beta_o + \beta_1 \underline{A} + \underline{v} \qquad (5.6)$$

then it can be found in, e.g., Fisz (1958, p. 89) that (even if $\underline{A}$ is stochastic)

$$\sigma_v^2 = \sigma_B^2(1-\rho^2) \qquad (5.7)$$

Remember that

$$\beta_1 = \rho\sigma_B/\sigma_A \qquad (5.8)$$

Substituting (5.7) and (5.8) into (5.4), it is simple to prove that (5.4) and (5.5) are identical.

Returning to eq. (5.4) we see that var $(\underline{\bar{B}})$ increases when the traffic intensity A can vary much from subrun to subrun, or when

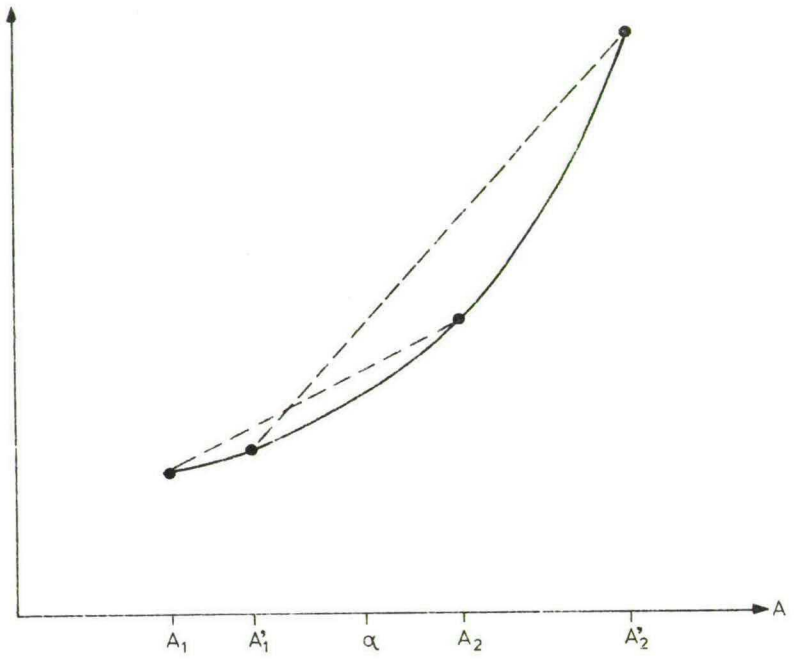

FIG. 5. Variation in $\beta_0 + \beta_1 A$ when $\vec{A}$ is replaced by $\vec{A}'$

the blocking probability B reacts strongly to changes in A (high β_1). This is an intuitively plausible result!

5.2. The regression estimator $\hat{B}_\alpha$

Applying eq. (2.1) yields

$$\text{var}\ (\underline{\hat{B}}_\alpha|\vec{A}) = \sigma_v^2 \{\frac{1}{n} + \frac{(\alpha-\bar{A})^2}{\Sigma(A_i-\bar{A})^2}\} \qquad (5.9)$$

Using (5.1) and (1.2) results in

$$\begin{aligned}\text{var}\ (\underline{\hat{B}}_\alpha) &= \underset{\vec{A}}{\mathcal{E}}\left[\sigma_v^2\{\frac{1}{n} + \frac{(\alpha-\bar{A})^2}{\Sigma(A_i-\bar{A})^2}\}\right] + \\ &+ \underset{\vec{A}}{\text{var}}\left[\mathcal{E}(\underline{\hat{\beta}}_0 + \underline{\hat{\beta}}_1.\alpha|\vec{A})\right] \\ &= \sigma_v^2/n + \sigma_v^2\mathcal{E}\{\frac{(\alpha-\underline{\bar{A}})^2}{\Sigma(\underline{A}_i-\underline{\bar{A}})^2}\} + \\ &+ \underset{\vec{A}}{\text{var}}\{\underline{\hat{\beta}}_0^{\vec{A}} + \underline{\hat{\beta}}_1^{\vec{A}}.\alpha\} \qquad (5.10)\end{aligned}$$

Let us try to compare this result to eq. (5.3). If $\alpha \approx \bar{A}$ then the second term of (5.10) will be small. In (5.3) $\underline{B}$ can move along the curve f(A). In (5.10) the variance term refers to changes in the regression parameters $\hat{\beta}_0$ and $\hat{\beta}_1$, as $\vec{A} = (A_1,\ldots,A_n)$ changes. See FIG. 5. More insight can be gained by approximating f(A) linearly.

If the true model f(A) is linear, then the regression model is corrected specified. In section 4.2 we proved that in that case the regression estimator $\underline{\hat{B}}_\alpha$ is unbiased for any $\vec{A}$. Consequently

$$\mathcal{E}(\underline{\hat{\beta}}_0 + \underline{\hat{\beta}}_1\,\alpha|\vec{A}) = \beta_0 + \beta_1\,\alpha \qquad (5.11)$$

Hence the variance component in (5.6) becomes zero. So

$$\text{var}\,(\hat{\underline{B}}_\alpha) = \sigma_v^2/n + \sigma_v^2 . \mathcal{E}\{\frac{(\alpha-\bar{\underline{A}})^2}{n.\underline{s}_A^2}\} \qquad (5.12)$$

where $\underline{s}_A^2$ is the estimated variance of $\underline{A}$. Comparing (5.12) and (5.4) shows that $\hat{\underline{B}}_\alpha$ has smaller variance, if

(i) the average $\bar{A}$ is close to α, or

(ii) the variance of $\underline{A}$ is large, or

(iii) the blocking probability reacts strongly to changes in A.

These are intuitively acceptable conclusions!

6. Conclusion and application

Above we proved that the simple, crude estimator $\bar{\underline{B}}$ is biased "ex post" (i.e., after the subrun traffic intensities A_i have been sampled), even if we assume that the blocking probability $\underline{B}$ reacts linearly to the traffic intensity $\underline{A}$ in the neighbourhood of the true input value $\alpha = \mathcal{E}(\underline{A})$. The regression estimator $\hat{\underline{B}}_\alpha$ is unbiased ex post (and hence also ex ante) assuming such a linear approximation. The variance of the regression estimator is also smaller than that of the crude estimator, if at least one of three intuitively interpretable conditions is satisfied. See also Table 1. Observe that in the crude estimator procedure the point estimator is based on the realized A_i-values, so that the bias is the ex post value $\beta_1(\bar{A}-\alpha)$. Concerning the variance, we have to realize that in a new experiment not only the $\underline{v}$ change, but also the $\underline{A}$. Hence the variance is based on the ex ante value. (In a new experiment the A-values change, and the crude estimator does not account for this effect.) Applying (5.5) and (5.7) to Table 1, it is easy to show that the MSE of the regression estimator becomes the smaller MSE, when ρ^2 approaches one (MSE = Mean Squared Error = variance plus squared bias). Notice that even if $\alpha = \bar{A}$, the crude estimator has higher MSE since the regression estimator can utilize the reactions of $\underline{B}$ to varying individual A_i.

	Crude $\bar{B}$	Regression $\hat{B}_\alpha$
Ex post (given $\vec{A}$)		
Bias	$\beta_1(\bar{A}-\alpha)$	0
Variance	σ_v^2/n	$\frac{\sigma_v^2}{n}\{1 + \frac{(\alpha-\bar{A})^2}{s_A^2}\}$
Ex ante (over all $\vec{A}$)		
Bias	0	0
Variance	$\frac{\sigma_B^2}{n} = \frac{\sigma_v^2}{n}\frac{1}{(1-\rho^2)}$	$\frac{\sigma_v^2}{n}\{1 + \frac{(\alpha-\bar{A})}{s_A^2}\}$

Table 1: Bias and variance of two estimators, assuming linear approximation near $A = \alpha$.

In our simulation experiment illustrated by FIG. 1 we obtained the following numerical results: $\hat{\sigma}_B = 0.364$ and $\hat{\sigma}_v = 0.179$

so that $\hat{\rho}^2 = 0.757$ or $\hat{\rho} = 0.870$. Since $\alpha = 28.700$, $\bar{A} = 28.819$ and $s_A = 0.425$, we can estimate the variance of $\hat{B}_\alpha$ as

$$\hat{\text{var}}\,(\hat{B}_\alpha) = \frac{\hat{\sigma}_v^2}{n}\{1 + \frac{(\alpha-\bar{A})^2}{s_A^2}\}$$

$$= \frac{0.032}{15}\{1 + \frac{(0.119)^2}{0.180}\} = 0.0023 \qquad (6.1)$$

so that its standard deviation is 0.048. Note that in this rather complicated simulation experiment, the service network was designed such that the blocking probability was hoped to equal 1%. Actually $\bar{B} = 1.545$ (in procents) with $\hat{\sigma} = 0.364$ so that $\hat{\sigma}_{\bar{B}} = 0.094$. Hence the simulation response is significantly worse that the desired value. For the regression estimator we obtain

$$\hat{B}_\alpha = \hat{\beta}_0 + \hat{\beta}_1 . \alpha =$$

$$= -20.198 + 0.754\ (28.7) = 1.4418 \qquad (6.2)$$

with its estimated standard deviation 0.048 so that this more accurate estimator also implies that the desired blocking probability is significantly exceeded. Note that narrow confidence intervals are desired since in future experiments different network structures will be compared.

The efficiency gain of regression sampling may be measured by the ratio var $(\bar{B})$/var $(\hat{B}_\alpha)$ or by its square root if the standard deviation is taken as criterion. When fixing the sample size n we are interested in the length of the confidence interval and hence in the standard deviation. When a prespecified confidence interval length ("accuracy") is desired, we are interested in the number of subruns needed to realize this accuracy. This number of subruns depends on the variance; see Kleijnen (1975). In the above example we have vâr $(\bar{B})$/vâr $(\hat{B}_\alpha) = 3.84$ and its square root is 1.960.

Observe that in the literature, including Kleijnen (1975), attention is focussed on the following regression or control variate; see also (1.3):

$$\hat{B}_\alpha = \bar{B} + \hat{\beta}_1 (\alpha - \bar{A}) \qquad (6.3)$$

If $\mathcal{E}(B|A) = f(A)$ can be approximated by a linear function in the neighbourhood of α, then we proved that (6.3) remains unbiased. In the general case this could not be proved. A "trick" discussed in Kleijnen (1975) is to use "jackknifing" to make $\hat{B}_\alpha$ unbiased and to derive confidence intervals.[7)]

NOTES

1) The unit of measurement is as follows. The number of arrivals per time unit is multiplied by the mean service time, i.e., the unit of time is the mean service time.

2) In eq. (1.5) σ_A is actually known since $\underline{A}$ is sampled from a known distribution. If we use the known σ_A in the estimator $\hat{\underline{\beta}}_c^*$, then (1.6) does not hold any more. However, it is convenient to apply the least squares algorithm. It may even be optimal to use (1.6) if errors in $\hat{\sigma}_B$ are compensated by errors in $\hat{\sigma}_A$; see Kleijnen (1975, p. 148).

3) In (2.1) we have $\vec{\beta}' = (\beta_0, \beta_1)$, $\vec{x} = (1, A_i)$, $\vec{y} = (\vec{B}_i)$.

4) In the textbooks on regression analysis the normality assumption is used to derive the distribution of $\hat{\underline{\beta}}$, which is needed to derive the standard t- and F- tests.

5) $\mathcal{E}(\hat{\underline{\vec{y}}} | \vec{x}' = \vec{x}_0') = \vec{x}'_0 \, \mathcal{E}(\hat{\vec{\beta}}) = \vec{x}_0' \cdot (\vec{X}'\vec{X})^{-1} \vec{X}' \mathcal{E}(\underline{\vec{y}})$

provided X is not stochastic (or if $\vec{X}$ is stochastic, $\vec{y}$ is not stochastically dependent on $\vec{X}$). Hence

$\mathcal{E}(\hat{\underline{\vec{y}}} | \vec{x}' = \vec{x}_0') = \vec{x}_0' \cdot (\vec{X}'\vec{X})^{-1} \vec{X}' (\vec{X}\beta) = \vec{x}_0' . \beta$

provided $\mathcal{E}(\vec{y}) = \vec{X} . \vec{\beta}$ is indeed a correct model.

6) In Kleynen (1975) the control variate estimator was derived to show bias in general and hence jackknifing was suggested). However, there $\mathcal{E}(\hat{\underline{\beta}})$ was standard whereas here $\mathcal{E}(\hat{\underline{\beta}} | A = \alpha)$ is of interest.

7) Consider $\hat{\underline{B}}_\alpha^{(i)} = \underline{\bar{B}} + \hat{\underline{\beta}}_1^{(i)} (\alpha - \underline{A}_i)$ where $\hat{\underline{\beta}}_1^{(i)}$ is the usual estimator of the slope β_1 but estimated from all pairs $(\underline{B}_j, \underline{A}_j)$ except pair $j = i$. Hence $\hat{\underline{\beta}}_1^{(i)}$ is independent of $\underline{A}_i$ so that $\mathcal{E}(\hat{\underline{\beta}}_1^{(i)} . \underline{A}_i) = \mathcal{E}(\hat{\underline{\beta}}_1^{(i)}) . \mathcal{E}(\underline{A}_i) = \alpha \, . \, \mathcal{E}(\hat{\underline{\beta}}_1^{(i)}$. Then it immediately follows that

$\mathcal{E}(\hat{\underline{B}}_{\alpha}^{(i)}) = \mathcal{E}(\underline{\bar{B}})$. So this "jackknifed" regression estimator has the same bias as the crude estimator. Its variance is, hopefully, smaller since it accounts for the reactions of $\underline{B}$ to deviations of $\underline{A}$ from α. See Kleijnen (1975) for more details.

REFERENCES

[1] BEAR, D., PRINCIPLES OF TELECOMMUNICATION - TRAFFIC ENGINEERING. P. Peregrinus Ltd., Southgate House, Stevenage (England), 1976.

[2] FISZ, M., WAHRSCHEINLICHKEITSRECHNUNG UND MATHEMATISCHE STATISTIK: V.E.B. Deutscher Verlag der Wissenschaften, Berlin, 1958. (English edition: PROBABILITY THEORY AND MATHEMATICAL STATISTICS. John Wiley & Sons, Inc., New York, third edition, 1967.

[3] JOHNSTON, J., ECONOMETRIC METHODS. McGraw-Hill Book Company, Inc., New York, 1963.

[4] KEEPING, E.S., INTRODUCTION TO STATISTICAL INFERENCE. D. Van Nostrand Company, Inc., Princeton, 1962.

[5] KLEIJNEN , J.P.C., STATISTICAL TECHNIQUES IN SIMULATION. (In two parts) Marcel Dekker, Inc., New York, 1974-1975.

[6] KLEIJNEN, J.P.C., Design and analysis of simulations: practical statistical techniques. SIMULATION, 28, no. 3, March 1977, pp. 81-90.

[7] SCHEFFÉ, H., THE ANALYSIS OF VARIANCE. John Wiley & Sons, Inc., New York, 4th printing, 1964.

PREVIOUS NUMBERS

EIT 54	J.J.M. Evers	A duality theory for convex ∞-horizon programming.
EIT 55	J. Dohmen J. Schoeber	Approximated fixed points.
EIT 56	J.J.M. Evers	Invariant competitive equilibrium in a dynamic economy with negotiable shares.
EIT 57	W.M. van den Goorbergh	Some calculations in a three-sector model.
EIT 58	W.G. van Hulst J.Th. van Lieshout	Investment/financial planning with endogeneous lifetimes: a heuristic approach to mixed-integer programming.
FEW 59	J.J.M. Evers M. Schubik	A dynamic economy with shares, fiat and accounting money.
FEW 60	J.M.G. Frijns	A dynamic model of factor demand equations
FEW 61	B.B. van der Genugten	A general approach to identification.
FEW 62	Pieter H.M. Ruys	Public goods and input-output analysis.
FEW 63	H.N. Weddepohl	Increasing returns and fixed market shares.
FEW 64	Claus Weddepohl	An equilibrium model with fixed labour time.
FEW 65	J.M.G. Frijns A. Hempenius	Dynamic optimal factor demand under financial constraints.
FEW 66	Jack P.C. Kleijnen	Operations research and computers.
FEW 67	Jack P.C. Kleijnen A.J. van den Burg R.Th. van der Ham	Generalizations of simulation results: Practicality of statistical methods (part 1).
FEW 68	A.J. van den Burg R.Th. van der Ham J.P.C. Kleijnen	Generalizations of simulation results: Practicality of statistical methods (part 2).
FEW 69	J. Pelkmans	Customs union under increasing costs conditions