NPS-AS-91-009

NAVAL POSTGRADUATE SCHOOL

Monterey, California

AN INVESTIGATION OF FINITE SAMPLE BEHAVIOR OF CONFIDENCE INTERVAL ESTIMATION PROCEDURES IN COMPUTER SIMULATION

Robert G. Sargent
Keebom Kang
David Goldsman

April 1991

Approved for public release; distribution unlimited.

Prepared for: Naval Postgraduate School
Monterey, CA 93943

NAVAL POSTGRADUATE SCHOOL
Monterey, California

RADM. R. W. West, Jr.
Superintendent

Harrison Shull
Provost

The research summarized herein was accomplished with resources provided by the Naval Postgraduate School.

This report was prepared by:

URITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

Form Approved
OMB No 0704-0188

REPORT SECURITY CLASSIFICATION Unclassfied	1b RESTRICTIVE MARKINGS
SECURITY CLASSIFICATION AUTHORITY	3 DISTRIBUTION/AVAILABILITY OF REPORT Approved for public release; distribution is unlimited
DECLASSIFICATION/DOWNGRADING SCHEDULE	
ERFORMING ORGANIZATION REPORT NUMBER(S) NPS-AS-91-009	5 MONITORING ORGANIZATION REPORT NUMBER(S) NPS-AS-91-009
NAME OF PERFORMING ORGANIZATION Naval Postgraduate School — 6b OFFICE SYMBOL (If applicable) AS	7a. NAME OF MONITORING ORGANIZATION Naval Postgraduate School
ADDRESS (City, State, and ZIP Code) Monterey, CA. 93943	7b. ADDRESS (City, State, and ZIP Code) Monterey, CA 93943
NAME OF FUNDING/SPONSORING ORGANIZATION Naval Postgraduate School — 8b OFFICE SYMBOL (If applicable)	9 PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER O&MN Direct Funding
ADDRESS (City, State, and ZIP Code) Monterey, CA 93943	10 SOURCE OF FUNDING NUMBERS PROGRAM ELEMENT NO — PROJECT NO — TASK NO — WORK UNIT ACCESSION NO

TITLE (Include Security Classification)
An Investigation of Finite Sample Behavior of Confidence Interval Estimation Procedures in Computer Simulation.

PERSONAL AUTHOR(S)
Robert G. Sargent, Keebom Kang and David Goldsman

TYPE OF REPORT	13b TIME COVERED	14 DATE OF REPORT (Year, Month, Day)	15 PAGE COUNT
Technical	FROM ____ TO ____	April 1991	35

SUPPLEMENTARY NOTATION

COSATI CODES			18 SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB-GROUP	Simulations, Statistical analysis; estimation, time series, confidence interval

ABSTRACT (Continue on reverse if necessary and identify by block number)

We investigate the small sample behavior and convergence properties of confidence interval estimators (CIE's) for the mean of a stationary discrete process. We consider CIE's arising from nonoverlapping batch means, overlapping batch means, and standardized time series, all of which are commonly used in discrete-event simulation. For a specific CIE, the performance measures of interest include the coverage probability, and the expected value and variance of the half-length. We use both empirical and analytical methods to make detailed comparisons regarding the behavior of the CIE's for a variety of stochastic processes. All of the CIE's under study are asymptotically valid; however, they are usually invalid for small sample sizes. We find that for small samples, the bias of the variance parameter estimator figures significantly in CIE coverage performance-the less bias the better. A Secondary role is played by the Marginal distribution of the stationary process. (continue on 2nd page)

DISTRIBUTION/AVAILABILITY OF ABSTRACT ☐ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT ☐ DTIC USERS	21. ABSTRACT SECURITY CLASSIFICATION Unclassified
NAME OF RESPONSIBLE INDIVIDUAL	22b. TELEPHONE (Include Area Code) (408) 646-3106 — 22c OFFICE SYMBOL AS/KK

orm 1473, JUN 86 — Previous editions are obsolete. — SECURITY CLASSIFICATION OF THIS PAGE

S/N 0102-LF-014-6603

continue from block 19. Abstract

We also point out that not all CIE's are equal - some require fewer observations before manifesting the properties for CIE Validity.

AN INVESTIGATION OF FINITE SAMPLE BEHAVIOR OF CONFIDENCE INTERVAL ESTIMATION PROCEDURES IN COMPUTER SIMULATION

Robert G. Sargent
Syracuse University
Syracuse, NY 13244

Keebom Kang
Naval Postgraduate School
Monterey, CA 93943

David Goldsman
Georgia Institute of Technology
Atlanta, GA 30332

ABSTRACT

We investigate the small sample behavior and convergence properties of confidence interval estimators (CIE's) for the mean of a stationary discrete process. We consider CIE's arising from nonoverlapping batch means, overlapping batch means, and standardized time series, all of which are commonly used in discrete-event simulation. For a specific CIE, the performance measures of interest include the coverage probability, and the expected value and variance of the half-length. We use both empirical and analytical methods to make detailed comparisons regarding the behavior of the CIE's for a variety of stochastic processes. All of the CIE's under study are asymptotically valid; however, they are usually invalid for small sample sizes. We find that for small samples, the bias of the variance parameter estimator figures significantly in CIE coverage performance - the less bias the better. A secondary role is played by the marginal distribution of the stationary process. We also point out that not all CIE's are equal - some require fewer observations before manifesting the properties for CIE validity.

Subject classifications: Simulation, statistical analysis, statistical estimation, time series, small sample behavior of confidence intervals

This paper studies the small sample behavior and convergence properties of a number of confidence interval estimators (CIE's) for the mean μ of a *stationary* process, $X_1,\ldots,X_n$. These CIE's are typically of the form

$$\Pr\{\mu \in \bar{\bar{X}}_n \pm t_{d,1-\alpha/2}(\hat{V}/n)^{1/2}\} \doteq 1 - \alpha, \tag{1}$$

where $\bar{\bar{X}}_n \equiv \Sigma_i X_i/n$, $t_{d,\gamma}$ is the γ-quantile of the t-distribution with d degrees of freedom (d.o.f.), and $\hat{V}$ estimates $\sigma_n^2 \equiv n\mathrm{Var}(\bar{\bar{X}}_n)$ (or the *variance parameter* $\sigma^2 \equiv \lim_{n\to\infty} \sigma_n^2$). A "good" estimator for σ^2 (or σ_n^2) is the cornerstone of a valid CIE for μ. Many estimators have been studied in the context of discrete-event simulation: nonoverlapping batch means (NOBM) (Conway 1963, Schmeiser 1982, Kang 1984); independent replications; overlapping batch means (OBM) (Meketon 1980, Meketon and Schmeiser 1984); standardized time series (STS) (Schruben 1983, Goldsman 1984, Glynn and Iglehart 1990); spectral theory (Fishman 1973,1978, Heidelberger and Welch 1981,1983); ARMA time series modeling (Fishman 1973,1978, Schriber and Andrews 1984); and regeneration (Crane and Iglehart 1975, Crane and Lemoine 1977, Fishman 1978).

There is considerable work which compares the various CIE methodologies. The Monte Carlo (MC) work mainly deals with small sample CIE performance; see, e.g., Law and Kelton (1984) and Goldsman, Kang, and Sargent (1986). Analytical results are almost all asymptotic: Goldsman and Schruben (1984), Goldsman and Meketon (1986), Damerdji (1987), Glynn and Iglehart (1990), and Schmeiser and Song (1989) all compare various combinations of the CIE's.

In this paper we study finite sample behavior of CIE's from NOBM, OBM, and STS. Section 1 gives background material. Section 2 reports on statistical properties of various variance estimators. Section 3 presents analytical results for some special cases and then summarizes a MC study of the CIE's. In Part I of our MC work, we break the n observations into b batches, and we observe what happens as the batch size grows. For "small" b, NOBM performs the best with respect to CIE *coverage*. For "large" b, both NOBM and OBM fare the best. However, an STS method produces CIE's with smaller expected lengths. Another comparison is carried out in Part II of our MC work, where we fix the d.o.f. d and observe what happens as n grows; here, the NOBM, OBM, and STS "combined" CIE's perform similarly. Section 4 discusses our findings, and Section 5 summarizes. We conclude that the bias of $\hat{V}$ is the most important factor in determining a CIE's validity; a secondary role is played by the marginal distribution of the X_i's. We also find that a CIE having superior large sample properties may have relatively poor small sample performance. We offer practical and research recommendations.

1. BACKGROUND

We review the CIE's and stochastic processes under study. We assume the stochastic processes satisfy certain moment and mixing conditions, as described in the cited papers. We henceforth use the notation $\mathrm{Nor}(\mu,\tau^2)$, $\chi^2(d)$, $\chi(d)$, $\mathrm{Exp}(\lambda)$, and $t(d)$ to represent the normal, chi-square, chi, exponential, and t distributions, each with the appropriate parameters.

1.1 Nonoverlapping Batch Means

Suppose we divide $X_1,\ldots,X_n$ into $b > 1$ adjacent, nonoverlapping batches of size m (assume $n = bm$). The i-th *batch mean*, $i = 1,\ldots,b$, is $\bar{X}_{i,m} \equiv \sum_{j=1}^{m} X_{(i-1)m+j}/m$. In implementing the method of NOBM, we assume the $\bar{X}_{i,m}$'s are approximately i.i.d. $\mathrm{Nor}(\mu,\sigma^2/m)$. The NOBM estimator for σ^2 is

$$\hat{V}_N \equiv m \sum_{i=1}^{b} [\, \bar{X}_{i,m} - \bar{\bar{X}}_n \,]^2/(b-1) \ \xrightarrow{\mathcal{D}}\ \sigma^2\chi^2(b-1)/(b-1),$$

where " $\xrightarrow{\mathcal{D}}$ " denotes convergence in distribution as $m \to \infty$. The NOBM CIE for μ

is given by (1) with $d = b-1$ and $\hat{V} = \hat{V}_N$.

1.2 Overlapping Batch Means

Define the i-th *overlapping batch mean*, $i = 1,\ldots,n-m+1$, by $\bar{X}(i,m) \equiv \sum_{j=0}^{m-1} X_{i+j}/m$. The OBM estimator for σ^2 is

$$\hat{V}_O \equiv nm \sum_{i=1}^{n-m+1} [\bar{X}(i,m) - \bar{\bar{X}}_n]^2/[(n-m+1)(n-m)] ,$$

and is almost identical to Bartlett's spectral estimator (see Priestley 1982). The OBM CIE for μ is given by (1) with $\hat{V} = \hat{V}_O$; its validity depends on $\hat{V}_O$ being approximately $\sigma^2\chi^2(d)/d$. Meketon and Schmeiser (1984) take $d = 1.5\cdot(b-1)$, where $b = n/m$. Based on MC experimentation, Schmeiser (1986) recommends $d = 1.5\cdot(b-1)[1+(b-1)^{-(.5+.6b)}]$; we shall use this value in our MC work.

1.3 Standardized Time Series

Suppose $X_1,\ldots,X_n$ is divided into $b \geq 1$ adjacent, nonoverlapping batches of size m. For $i = 1,\ldots,b$, let $\hat{A}_i \equiv \sum_{j=1}^{m} [(m+1)/2 - j]X_{(i-1)m+j}$. Schruben (1983) assumes the $\hat{A}_i$'s are approximately i.i.d. normal, and proposes the *area* and *combined* estimators for σ^2:

$$\hat{V}_A \equiv \frac{12}{(m^3-m)b} \sum_{i=1}^{b} \hat{A}_i^2 \xrightarrow{\mathfrak{D}} \frac{\sigma^2\chi^2(b)}{b} \quad \text{and} \quad \hat{V}_C \equiv \frac{(b-1)\hat{V}_N + b\hat{V}_A}{2b-1} \xrightarrow{\mathfrak{D}} \frac{\sigma^2\chi^2(2b-1)}{2b-1} .$$

CIE's for μ are formed by substituting the appropriate $\hat{V}$ and d in (1). Schruben also derives the so-called "maximum" estimator for σ^2 (see Subsection 4.2).

1.4 Some Time Series Processes of Interest

We will have occasion to use the following ARMA-type processes.

MA(1): $X_i = \varepsilon_i + \theta\varepsilon_{i-1}$, where $\varepsilon_i \sim$ i.i.d. Nor(0,1) and $-1 < \theta < 1$.

MA'(1): $X_i = \varepsilon_i + \theta\varepsilon_{i-1}$, where $\varepsilon_i \sim$ i.i.d. Exp(1) and $-1 < \theta < 1$.

AR(1): $X_i = \varphi X_{i-1} + \varepsilon_i$, where $\varepsilon_i \sim$ i.i.d. Nor$(0,1-\varphi^2)$, $-1 < \varphi < 1$.

$$\text{EAR}(1):\ X_i = \begin{cases} \varphi X_{i-1} & \text{w.p. } \varphi \\ \varphi X_{i-1} + \varepsilon_i & \text{w.p. } 1-\varphi \end{cases}, \quad \text{where } \varepsilon_i \sim \text{i.i.d. Exp}(1) \text{ and } 0 \le \varphi < 1 \text{ (see Lewis 1980).}$$

We also consider other stationary normal processes as well as the waiting time (delay) processes of M/M/1 and E_r/M/1 queues.

1.5 CIE Performance Criteria

Denote the NOBM, OBM, area, and combined CIE's by CIE_N, CIE_O, CIE_A, and CIE_C, resp. The *half-length* of a generic CIE is $H \equiv t_{d,1-\alpha/2}(\hat{V}/n)^{1/2}$. We use the following CIE performance measures: coverage ($\Pr\{\mu \in \bar{\bar{X}}_n \pm H\}$), $E[H]$, and $\text{Var}(H)$. Among CIE's which achieve coverage $1 - \alpha$, we prefer that with the smallest $E[H]$, and then that with the smallest $\text{Var}(H)$.

2. PROPERTIES OF VARIOUS VARIANCE ESTIMATORS

We give some results on the bias and variance of the NOBM, OBM, and STS estimators, and on the asymptotic performance of the corresponding CIE's.

2.1 Bias of the Estimators

The bias of $\hat{V}$ as an estimator for σ^2 is $\text{Bias}(\hat{V}) \equiv \sigma^2 - E[\hat{V}]$. Goldsman and Meketon (1986) show that $\lim_{m\to\infty} m \lim_{b\to\infty} \text{Bias}(\hat{V}_N) = \lim_{m\to\infty} m \lim_{b\to\infty} \text{Bias}(\hat{V}_O) = \lim_{m\to\infty} m \lim_{b\to\infty} \text{Bias}(\hat{V}_C)/2 = \lim_{m\to\infty} m\, \text{Bias}(\hat{V}_A)/3$. All of these estimators are asymptotically unbiased as $m \to \infty$, but for small m, these estimators can be quite biased.

<u>Example 1</u>: For the AR(1) and EAR(1), the Appendix gives $\sigma^2 = (1+\varphi)/(1-\varphi)$ and

$$E[\hat{V}_N] = \sigma^2 - \frac{2\varphi(b+1)}{mb(1-\varphi)^2} + \frac{2\varphi b(\varphi^m - b^{-2}\varphi^{mb})}{m(b-1)(1-\varphi)^2} \doteq \sigma^2 - \frac{2\varphi}{m(1-\varphi)^2} \text{ for large m and b,} \qquad (2)$$

$$E[\hat{V}_O] = \sigma^2 - \frac{2b\varphi}{m(b-1)(1-\varphi)^2}\left[1 - \varphi^m + \frac{1-\varphi^{mb}}{b^2}\right] + \frac{4\varphi(1-\varphi^m)(1-\varphi^{mb-m+1})}{m(b-1)(mb-m+1)(1-\varphi)^3} \tag{3}$$

$$\doteq \sigma^2 - \frac{2\varphi}{m(1-\varphi)^2} \text{ for large m and b,}$$

$$E[\hat{V}_A] = \sigma^2 + \frac{24\varphi}{(m^3-m)(1-\varphi)^2}\left\{\frac{-m^2+1-\varphi^m(m+1)^2}{4} + \frac{\varphi-(m+1-m\varphi)\varphi^{m+1}}{(1-\varphi)^2}\right\} \doteq \sigma^2 - \frac{6\varphi}{m(1-\varphi)^2} \tag{4}$$

for large m,

$$E[\hat{V}_C] = \sigma^2 + \frac{2\varphi}{(2b-1)m(1-\varphi)^2}\left\{-4b + \frac{1-\varphi^{mb}}{b} - \frac{2b(m+2)\varphi^m}{m-1} + \frac{12b\varphi[1-(m+1-m\varphi)\varphi^m]}{(m^2-1)(1-\varphi)^2}\right\} \tag{5}$$

$$\doteq \sigma^2 - \frac{4\varphi}{m(1-\varphi)^2} \text{ for large m and b.}$$

If m and b are large, the bias results anticipated by Goldsman and Meketon are attained. Table 1 contains exact $E[\hat{V}]$'s for b = 2 and 16 and various m. For small b, $\text{Bias}(\hat{V}_N) < \text{Bias}(\hat{V}_O) < \text{Bias}(\hat{V}_C) < \text{Bias}(\hat{V}_A)$; for large b, $\text{Bias}(\hat{V}_N) \doteq \text{Bias}(\hat{V}_O) < \text{Bias}(\hat{V}_C) < \text{Bias}(\hat{V}_A)$. The biases become negligible as m grows. //

<u>Example 2</u>: For the MA(1) and MA'(1), the Appendix gives $\sigma^2 = (1+\theta)^2$ and

$$E[\hat{V}_N] = \sigma^2 - 2\theta(b+1)/mb \doteq \sigma^2 - 2\theta/m \quad \text{for large b.} \tag{6}$$

$$E[\hat{V}_O] = \sigma^2 - \frac{2\theta}{m(b-1)}\left[\frac{b^2+1}{b} - \frac{2}{mb-m+1}\right] \doteq \sigma^2 - 2\theta/m \quad \text{for large b.} \tag{7}$$

$$E[\hat{V}_A] = \sigma^2 - 6\theta/m. \tag{8}$$

$$E[\hat{V}_C] = \sigma^2 - (4b+2)\theta/mb \doteq \sigma^2 - 4\theta/m \quad \text{for large b.} \tag{9}$$

The conclusions from Example 1 again hold. //

Although $\text{Bias}(\hat{V})$ is interesting in its own right, the bias directly affects CIE performance. Consider the unrealistic case in which m is fixed and $b \to \infty$. For the estimators studied here, one can show that as $b \to \infty$, $\hat{V} \to E[\hat{V}]$ w.p.1, and so $T \equiv (\bar{\bar{X}}_n-\mu)/(\hat{V}/n)^{1/2} \xrightarrow{\mathcal{D}} \text{Nor}(0,\sigma^2/E[\hat{V}])$. Falsely assuming T ~ Nor(0,1) as $b \to \infty$ yields incorrect CIE coverage,

$$\Pr\{\mu \in \bar{\bar{X}}_n \pm H\} \to 2\Phi[z_{1-\alpha/2}(E[\hat{V}]/\sigma^2)^{1/2}] - 1 \;, \tag{10}$$

where $\Phi(\cdot)$ and z_γ are the Nor(0,1) c.d.f. and γ-quantile. If $E[\hat{V}]/\sigma^2 < [>]\ 1$, then coverage $< [>]\ 1 - \alpha$. (Coverage is quite sensitive to decreases in $E[\hat{V}]/\sigma^2$.) So the less bias the better. As $b \to \infty$, CIE_N and CIE_O tend to achieve the desired coverage more quickly with respect to m than do CIE_A and CIE_C; see Sargent, Kang, and Goldsman (1987) (S-K-G).

2.2 Variance of the Estimators

Goldsman and Meketon report that as m and b become large, $b\cdot\mathrm{Var}(\hat{V}_N) \to 2\sigma^4$, $b\cdot\mathrm{Var}(\hat{V}_O) \to 4\sigma^4/3$, $b\cdot\mathrm{Var}(\hat{V}_A) \to 2\sigma^4$, and $b\cdot\mathrm{Var}(\hat{V}_C) \to \sigma^4$ (cf. Damerdji 1987). For i.i.d. $X_1,\ldots,X_n$, Kang and Goldsman (1990) find $\mathrm{Var}(\hat{V}_N)$ and $\mathrm{Var}(\hat{V}_A)$ exactly, as do Song and Schmeiser (1989) for $\mathrm{Var}(\hat{V}_O)$. Exact results for other processes and for $\mathrm{Var}(\hat{V}_C)$ are tedious to derive. Of course, one can also calculate the mean squared errors of the $\hat{V}$'s (cf. Schmeiser and Song 1989).

2.3 Asymptotic Properties of the CIE's

Schmeiser (1982) and Goldsman and Schruben (1984) note that as $m \to \infty$,

$$(mb)^{1/2}H \xrightarrow{\mathcal{D}} \sigma t_{d,1-\alpha/2}\chi(d)/\sqrt{d} \tag{11}$$

for CIE_N, CIE_A, and CIE_C. An analogous *approximate* result holds for CIE_O. Under (11), the CIE's achieve coverage $1 - \alpha$ as $m \to \infty$. Further, if $\Gamma(\cdot)$ is the gamma function, then

$$(mb)^{1/2}E[H] \to \sigma t_{d,1-\alpha/2}(2/d)^{1/2}\,\frac{\Gamma((d+1)/2)}{\Gamma(d/2)}\;,\ \text{and} \tag{12}$$

$$mb\mathrm{Var}(H) \to \sigma^2 t^2_{d,1-\alpha/2}\left\{1 - \frac{2}{d}\left[\frac{\Gamma((d+1)/2)}{\Gamma(d/2)}\right]^2\right\}. \tag{13}$$

The right sides of (12) and (13) *decrease in d* and, hence, in b. So for large m with fixed b, $E[H_N] > E[H_A] > E[H_O] > E[H_C]$ and $\mathrm{Var}(H_N) > \mathrm{Var}(H_A) > \mathrm{Var}(H_O) > \mathrm{Var}(H_C)$, the subscripts having the obvious meanings. Goldsman and Schruben

(1984) and Meketon and Schmeiser (1984) let $b \to \infty$ in (12) and (13) to get

$$\lim_{\substack{b\to\infty \\ m\to\infty}} \frac{E[H]}{E[H_N]} = 1,\ H = H_O, H_A, H_C, \quad \text{and} \quad \lim_{\substack{b\to\infty \\ m\to\infty}} \frac{\mathrm{Var}(H)}{\mathrm{Var}(H_N)} = \begin{cases} 1, & H = H_A \\ 2/3, & H = H_O \\ 1/2, & H = H_C \end{cases}. \tag{14}$$

Thus, as b also becomes large, all of the CIE's have about the same E[H]'s, but CIE_C has the smallest Var(H).

3. FINITE SAMPLE CONFIDENCE INTERVAL ESTIMATION

Small sample analysis of CIE's is difficult. We present a few exact results, but most of the section is devoted to a MC study.

3.1 Some Analytical Results

Example 3: Suppose $X_1,\ldots,X_n \sim$ i.i.d. Nor(μ,τ^2). Then $\hat{V}_N \sim \tau^2\chi^2(b-1)/(b-1)$, $\hat{V}_A \sim \tau^2\chi^2(b)/b$, and $\hat{V}_C \sim \tau^2\chi^2(2b-1)/(2b-1)$. Further, $\bar{\bar{X}}_n$ is independent of $\hat{V}_N$, $\hat{V}_A$, and $\hat{V}_C$ (see Appendix). So (1) holds exactly for CIE_N, CIE_A, and CIE_C. We could not obtain such results for CIE_O or for nonnormal i.i.d. processes. //

Example 4: We can derive exact results for CIE_A when b = 1 (n = m) and $X_1,\ldots,X_n$ is *stationary normal*. Then $\hat{A}_1 \sim$ Nor$(0,E[\hat{A}_1^2])$, $\hat{V}_A \sim E[\hat{V}_A]\chi^2(1)$, and $\bar{\bar{X}}_n$ is normal and independent of $\hat{V}_A$ (see Appendix). So $(\bar{\bar{X}}_n-\mu)(nc/\hat{V}_A)^{1/2} \sim t(1)$, where $c \equiv E[\hat{V}_A]/\sigma_n^2$. Hence, the coverage of CIE_A is $2\Pr\{t(1) \le t_{1,1-\alpha/2}\sqrt{c}\} - 1 = (2/\pi)\mathrm{Tan}^{-1}(t_{1,1-\alpha/2}\sqrt{c})$. (As in (10), the coverage is sensitive to decreases in c.) Similarly, $E[H_A] = t_{1,1-\alpha/2}(2E[\hat{V}_A]/\pi n)^{1/2}$. To illustrate, suppose the X_i's are AR(1). Figure 1(a) uses Example 1 and (A-3) to plot coverage vs. $\log_2 n$ for $1 - \alpha = 0.90$ and various φ. We see that for $\varphi > [<]\ 0$, the coverage is $< [>]\ 1 - \alpha$. As $|\varphi|$ approaches 0 or as n grows, Bias$(\hat{V}_A)$ decreases, and the coverage approaches $1 - \alpha$. Figure 1(b) has analogous plots of $E[H_A]$ vs.

$\log_2 n$. If $\varphi = 0.0$, then $E[\hat{V}_A] = \sigma^2 = 1$, and so $E[H_A]$ decreases at rate $n^{-½}$. If $\varphi = -0.9$, then $E[\hat{V}_A]$ *decreases* to $\sigma^2 = 1/19$, and $E[H_A]$ decreases at rate $n^{-½}$ (after initially decreasing faster). The $\varphi = 0.9$ plot for $E[H_A]$ *increases and then decreases* as n grows. This occurs since $E[\hat{V}_A]$ increases to $\sigma^2 = 19$ as $n \to \infty$, while the $\sqrt{n}$ in $E[H_A]$'s denominator becomes large. //

Example 5: We give exact results for CIE_N when $b = 2$ $(n = 2m)$ and $X_1,\ldots,X_n$ is stationary normal. Then $\hat{V}_N = m\Sigma_i(\bar{X}_{i,m} - \bar{\bar{X}}_n)^2/(b-1) = m(\bar{X}_{1,m} - \bar{X}_{2,m})^2/2$; so $\hat{V}_N \sim E[\hat{V}_N]\chi^2(1)$. Since $\bar{\bar{X}}_n$ is normal and independent of $\hat{V}_N$ for $b = 2$ (see Appendix), we have $(\bar{\bar{X}}_n - \mu)(nc'/\hat{V}_N)^{½} \sim t(1)$, where $c' \equiv E[\hat{V}_N]/\sigma_n^2$. The coverage is $(2/\pi)\mathrm{Tan}^{-1}(t_{1,1-\alpha/2}\sqrt{c'})$, and $E[H_N] = t_{1,1-\alpha/2}(E[\hat{V}_N]/\pi m)^{½}$. If the X_i's are AR(1), then (A-3) yields σ_n^2, and (2) with $b = 2$ gives $E[\hat{V}_N]$; c', coverage, and $E[H_N]$ then follow. These performance measures behave as in Example 4. //

It is difficult to generalize the above CIE results to $b > 2$, since we would then have *correlated* χ^2 random variables. Thus, we only gave exact results for simple cases. We resort to MC experimentation in the sequel.

3.2 Design of the Monte Carlo Study

Our goal was to assess CIE performance over a variety of stochastic processes and choices of number of observations n, batch size m, d.o.f. d, and desired coverage $1 - \alpha$. We simulated the following processes: AR(1) with $\varphi = 0.0, \pm0.1, \pm0.5, \pm0.9$; EAR(1) with $\varphi = 0.0, 0.1, 0.5, 0.9$; MA(1) and MA'(1) with $\theta = \pm0.1, \pm0.9$; M/M/1 with traffic intensity $\rho = 0.6, 0.8$ (service rate = 1.0); and E_3/M/1 with $\rho = 0.6$ (service rate = 1.0). Each run was initialized from the appropriate *steady state* distribution. All uniform [normal] variates were generated from algorithm UNIF [TRPNRM] in Bratley, Fox, and Schrage (1987); exponentials used inversion.

For ease of exposition, we divided the study into two parts. In Part I we fixed $1 - \alpha$ and b, and then charted CIE coverage as a function of m. Roughly speaking, we wanted to know which CIE first achieved acceptable coverage as m increased with fixed b. Further, which CIE's had the best E[H] and (to a lesser extent) Var(H)? For the stochastic processes discussed above, we conducted at least 1000 independent runs of 16384 observations; these runs were used to calculate CIE_N, CIE_O, CIE_A, and CIE_C and the resulting performance characteristics for all choices of $m = 2^k$, $k = 0,1,\ldots,10$, $b = 1,2,4,8,16$, and $1 - \alpha = 0.80, 0.90, 0.95, 0.99$.

In Part II, we set $1 - \alpha = 0.90$ and $d = 3$ and 15, and then charted coverage as a function of n. For $d = 3$ [$d = 15$], CIE_N uses $b = 4$ [$b = 16$], CIE_O uses $b = 2$ [$b = 11$], CIE_A uses $b = 3$ [$b = 15$], and CIE_C uses $b = 2$ [$b = 8$]. We conducted our experiments on a number of stochastic processes, each of which used 2000 runs of 16384 observations to calculate the CIE's for all $n = 2^k$, $k = 4,\ldots,14$, and $d = 3$ and 15. Since d is fixed, (11)-(13) suggest that all of the CIE's will perform about the same for large m; however, we suspected that they would exhibit different small sample performance because the variance estimators incorporated in the CIE's operate under different assumptions. We first discuss the underlying assumptions for $\hat{V}_N$, $\hat{V}_A$, and $\hat{V}_C$ since these estimators require independence between batches ($\hat{V}_O$ does not). The estimator $\hat{V}_N$ [$\hat{V}_A$] assumes that the $\bar{X}_{i,m}$'s [$\hat{A}_i$'s] are i.i.d. normal; the combined estimator $\hat{V}_C$ must satisfy both assumptions. For fixed d and n, $\hat{V}_N$ and $\hat{V}_A$ use roughly half the batch size of $\hat{V}_C$; hence, the assumption of i.i.d. $\bar{X}_{i,m}$'s [$\hat{A}_i$'s] is harder to achieve for $\hat{V}_N$ [$\hat{V}_A$] than for $\hat{V}_C$. On the other hand, $\hat{V}_N$'s assumption of normality of the $\bar{X}_{i,m}$'s is easier to satisfy than $\hat{V}_A$'s assumption of normality of the $\hat{A}_i$'s which relies on a more restrictive functional central limit theorem. The normality question for $\hat{V}_N$ vs. $\hat{V}_C$ is not as clear since $\hat{V}_N$ uses half of $\hat{V}_C$'s batch size. CIE_0 appeals to spectral theory to directly assume that $\hat{V}_0$ is $\sigma^2\chi^2(d)/d$; this supposition is not true for nonnormal processes or for finite batch sizes.

3.3 Results from Part I of the Monte Carlo Study

3.3.1 Representative results

We discuss *typical* results from Part I. Figures 2, 3, and 4 illustrate CIE performance when $X_1,\ldots,X_n$ are AR(1) (φ = 0.9), EAR(1) (φ = 0.9), and M/M/1 (ρ = 0.8), resp., and $1 - \alpha = 0.90$. The AR(1) and EAR(1) have the same covariance function, but the AR(1) has normal marginals while the EAR(1)'s are exponential. The M/M/1's joint distribution is more complicated. We only consider the cases b = 2 and 16 since CIE_N, CIE_0, and CIE_C require $b \geq 2$, and since one can argue that b = 16 is "large" (at least for the "usual" choices of α). Each of Figures 2, 3, and 4 has four graphs: (a),(b) are for b = 2, and (c),(d) are for b = 16. In (a) and (c), we plot the sample coverage (CVG) vs. $\log_2 m$; (b) and (d) plot the sample E[H] (EHL) vs. $\log_2 m$. The standard error of any CVG is about $[CVG\cdot(1-CVG)/1000]^{1/2}$.

Figure 2 is for the AR(1) with φ = 0.9. All CVG's are poor for small m, but approach $1 - \alpha = 0.90$ as m increases. For b = 2, the CVG of CIE_N is closest to 0.90; for b = 16, both CIE_N and CIE_0 yield the best CVG's. This makes sense since $\hat{V}_N$ (for b = 2 and 16) and $\hat{V}_0$ (for b = 16) are less biased than $\hat{V}_A$ and $\hat{V}_C$ (Example 1 and Table 1). A related consequence is that CIE_N and CIE_0 produce larger EHL's than those of CIE_C. Note that the EHL's in Figure 2 *increase and then decrease* as m increases - the same bias-related pattern as for the exact $E[H_A]$ in Figure 1. The EHL's for b = 16 and $m \geq 128$ from Figure 2(d) are more or less in agreement with the limiting $E[H]/E[H_N] = 1$ ratios in (14); this is one reason why we regard b = 16 as "large."

Table 2 has CVG's for the AR(1) with $b = 16$, $1 - \alpha = 0.90$, and various m and φ. These roughly agree with the $b \to \infty$ results from (10). We see that CVG < [>] 0.90 when φ > [<] 0.0. For any φ and fixed m, Table 2 shows that the CVG's of CIE_N and CIE_O are closer to $1 - \alpha$ than those of CIE_A and CIE_C.

Results for the EAR(1) process with $\varphi = 0.9$ are found in Figure 3, whose plots bear a striking resemblance to those of the AR(1) in Figure 2. The only notable difference between the two figures is that, for fixed m, the EAR(1) has smaller CVG's (see Table 1). Since the AR(1) and EAR(1) have the same covariance structure, the EAR(1)'s poorer CVG's are probably due to its exponential marginals. It seems that the effect of Bias($\hat{V}$) on coverage is more significant than that of the marginals of the X_i's.

Figure 4 concerns the M/M/1 process with $\rho = 0.8$. Again, the patterns in the figure are not much different than those of the AR(1) and EAR(1). The M/M/1 simply requires more observations to attain valid coverage. The positive serial correlation of the M/M/1 causes the estimators for σ^2 to be biased too low; so poor coverage results for small m.

3.3.2 Additional Part I results

In S-K-G, we also give detailed discussions on (among other things):

- CIE performance for the MA(1) as θ varies. We find that the CIE's qualitatively perform about the same as those for the AR(1).

- The sample $\sqrt{Var(H)}$ (SHL) performance measure. For small m, the SHL's exhibit the same general behavior as the EHL's; as m and b become large, the SHL's behave as in (14). Even though the ratios from (14) are manifested, the *differences* between SHL's from competing methods are typically very small.

- The consequences of changing $1 - \alpha$. The CIE's qualitatively perform the same as we vary $1 - \alpha$. Coverage is sensitive to decreases in $1 - \alpha$. For example, consider CIE_N for the AR(1) with $\varphi = 0.9$, $b = 2$, and $m = 16$; for $1 - \alpha = 0.99$, 0.90, and 0.80, Example 5 gives coverages of 0.985, 0.852, and 0.713, resp.

3.3.3 Summary of Part I

For small m, improper CVG's were usually the rule. For small m *and* b, CIE_N has better CVG's than the other CIE's, while for small m and large b, both CIE_N and CIE_O fared the best. For fixed m, high CVG was most often accompanied by high EHL and SHL. The CIE's performed as expected by asymptotic theory when m and/or b became large. For large m and small b, all achieved the desired coverage, and the EHL's and SHL's tended to decrease with increasing d.o.f., as per Subsection 2.3. For large m *and* b, the ratios from (14) took effect.

3.4 Results from Part II of the Monte Carlo Study

We considered the MA(1) ($\theta = -0.9$), AR(1) ($\varphi = 0.9$), EAR(1) ($\varphi = 0.9$), and M/M/1 ($\rho = 0.8$), with $1 - \alpha = 0.90$ and d.o.f. $d = 3$ and 15. Table 3 gives CVG's as a function of the sample size n. For $d = 3$, CIE_N, CIE_O, and CIE_C perform about the same in terms of CVG; CIE_A fares poorly for small n. For $d = 15$ and small n, CIE_O does a bit better than the others in terms of CVG; CIE_A is not competitive. However, the performance of CIE_O is "more variable" than the other CIE's over the range of stochastic processes, d, and n under study. For instance, the CVG of CIE_O sometimes significantly overshoots $1 - \alpha$, especially for small d (though this is understandable since OBM was designed for large b). As n grows (for $d = 3$ or 15), it appears that CIE_N, CIE_O, and CIE_C achieve CVG $\doteq 0.90$ at about the same n.

4. DISCUSSION

We first discuss the causes of improper CIE coverage; why are some CIE's better than others? We then consider the question of which CIE is "best?"

4.1 Causes of Improper Coverage

A CIE of the form in (1) will attain perfect coverage if its associated pivot $T \equiv (\bar{\bar{X}}_n - \mu)/(\hat{V}/n)^{1/2} \sim t(d)$; this requires (i) $\bar{\bar{X}}_n \sim \mathrm{Nor}(\mu, \sigma_n^2/n)$, (ii) $\hat{V} \sim \sigma_n^2 \chi^2(d)/d$, and (iii) independence of $\bar{\bar{X}}_n$ and $\hat{V}$. Requirement (i) is satisfied if the marginal distribution of the X_i's is normal. If the X_i's are not symmetric, then T may be skewed for small n. But in most cases, a central limit theorem asserts that (i) approximately holds as n grows. We believe that (ii) is the key requirement. At a minimum, $\hat{V}$ must be approximately unbiased as an estimator of σ^2 (or σ_n^2). In fact, since variance directly affects the CIE's length, we claim that Bias($\hat{V}$) is often the main cause of improper coverage (at least for small m); see below. Concerning (iii), Glynn (1982) and Kang and Goldsman (1990) demonstrate that *asymmetry* in coverage is directly related to dependence between $\bar{\bar{X}}_n$ and $\hat{V}$. However, Kang and Goldsman give examples which show that actual coverage is not necessarily affected by such dependence. We do not view dependence between $\bar{\bar{X}}_n$ and $\hat{V}$ as a direct cause of improper coverage.

We first analyze the effect of Bias($\hat{V}$) on requirement (ii) by examining $\tilde{V} \equiv \sigma_n^2 \hat{V}/E[\hat{V}]$ instead of $\hat{V}$; note that $E[\tilde{V}] = \sigma_n^2$. To illustrate, we shall use the NOBM estimator on the AR(1) and EAR(1) processes with $\varphi = 0.9$. For the AR(1) with b = 2, Example 5 says that $\hat{V}_N \sim E[\hat{V}_N]\chi^2(1)$, and so $\tilde{V}_N \sim \sigma_n^2\chi^2(1)$; for this case, correction for bias results in precisely the desired distribution. Empirical p.d.f.'s of $\hat{V}_N/\sigma_n^2$ and $\tilde{V}_N/\sigma_n^2$ (based on 100000 independent runs) are plotted in Figure 5 for the AR(1) and EAR(1) with b = 8 and various m. For the AR(1), the sample p.d.f.'s of $\hat{V}_N/\sigma_n^2$ approach the $\chi^2(7)$ p.d.f. as m increases;

the corrected $\tilde{V}_N/\sigma_n^2$ is nearly (but not quite) $\chi^2(7)$ (Figures 5(a) and 5(b)). The empirical p.d.f.'s for the EAR(1) exhibit similar behavior, except that the convergence to the $\chi^2(7)$ is slower (Figures 5(c) and 5(d)). Thus, for these examples, correction for bias mitigates the violation of (ii). Comparing the AR(1) results with the EAR(1)'s, we conclude that the effect of Bias($\hat{V}$) is greater than that of the marginal distribution.

We will next show for the stochastic processes investigated that correcting for Bias($\hat{V}$) results in good CIE's. The notation CCVG in Table 1 is the exact (from Example 5) or sample coverage obtained for the AR(1) and EAR(1) models with $\varphi = 0.9$ and $b = 2$ and 16. For the AR(1) with $b = 2$ and any m, the corrected NOBM pivot $(\bar{\bar{X}}_n - \mu)/(\tilde{V}_N/n)^{1/2} \sim t(1)$. So the CCVG's for CIE_N are *exactly* $1 - \alpha = 0.90$; thus, for this example, Bias($\hat{V}_N$) is the sole cause of improper coverage. The corresponding MC CCVG improvements for the EAR(1) with $b = 2$ are significant but not as good as those for the AR(1), this indicating that a secondary marginal distributional effect is present. This conclusion is illustrated yet again by Figure 6. (The EAR(1)'s empirical p.d.f.'s are somewhat skewed for small m as explained earlier.) The results from Table 1 and Figures 5 and 6 suggest that, for small m, there are also tertiary contributors to improper coverage, perhaps inter-batch correlation.

4.2 Which is the Best CIE?

The question of which CIE is the "best" depends on the criteria being used. A CIE is first judged by the validity of its coverage. But the MC work showed that a CIE with good coverage might produce relatively wide half-lengths; so the E[H] and Var(H) measures can not be ignored. Since coverage, E[H], and Var(H) are functions of the stochastic process, the CIE in use, the number of batches b, the batch size m, and the level α, we can see that the determination of the "best" CIE is not straightforward.

Results (12) and (13) say that E[H] and Var(H) decrease in the d.o.f. d as m becomes large with fixed b; so the more d.o.f. the better (although Schmeiser 1982 finds that there is little to be gained by taking $d > 30$). Nevertheless, it would be incorrect to conclude that CIE_C (which has the largest d.o.f. of those CIE's under study) is always the best. The Part I MC results showed that for small b, CIE_N required the smallest value of m to achieve valid coverage (for large b, CIE_N and CIE_O required the smallest m); if coverage were the only criterion for CIE comparison, CIE_N would be declared the best - not CIE_C. This shows that *large sample superiority does not necessarily extend to the small sample case*. Indeed, we did not include the STS "NOBM+maximum" CIE from Schruben (1983), which has 4b-1 d.o.f. (and hence superior asymptotic properties), since it exhibited poor small sample performance compared to the other CIE's (including CIE_A). For instance, for the AR(1) with $\varphi = 0.9$, $1 - \alpha = 0.90$, $b = 2$, and $m = 16$, 64, and 256, the NOBM+max CIE attained CVG's (based on 1000 independent simulation runs) of 0.397, 0.606, and 0.776, resp.; CIE_A achieved CVG's of 0.684, 0.849, and 0.895, resp. (Table 1).

Another basis for comparison among CIE's is to determine which requires the smallest sample size n to achieve valid coverage for some fixed d.o.f. d. This was the aim of Part II of the MC study, where CIE_N, CIE_O, and CIE_C fared more or less the same; CIE_A was not competitive. So there was no clear winner using the criterion of coverage under fixed d.o.f.

We can still make some recommendations (for fixed sample size procedures). All of the CIE's studied here are easy to use. Batch means is the simplest method to understand. All are asymptotically valid as the batch size m grows; but it "never hurts too much" to use CIE_N (in comparison to the other CIE's) in case m is not "large enough." The price to be paid when m is small is that $E[H_N]$ and $Var(H_N)$ are larger than their competitors, particularly when b is also small. If the user is somehow confident that the batch sizes are large enough to achieve valid coverage, then the user should fix the d.o.f. (perhaps between 15 and 30 for the "usual" α values) and select from among CIE_N, CIE_O, and CIE_C. However, one of the most difficult open questions in simulation output analysis is the determination of "sufficient batch size" (cf. Fishman 1978 and Schmeiser and Song 1989).

5. SUMMARY AND CONCLUSIONS

We studied the behavior of different CIE's with special emphasis placed on small sample size performance for various stationary stochastic processes. *If the achieved coverages are at the desired levels*, and if the total number of observations n is fixed, the ranking of the CIE's with respect to E[H] and Var(H) is determined by the d.o.f. each CIE has. In this case, the CIE with the largest d.o.f. has the smallest mean and most stable confidence interval length.

Perhaps our most important finding was that, in small sample settings, a CIE with more d.o.f. may not actually be "better" than a competing CIE; some CIE's may require more observations than others before the asymptotics necessary for CIE validity manifest themselves. Quite often, the CIE's with the highest d.o.f.'s performed the most poorly in terms of coverage!

The bias of $\hat{V}$ as an estimator of σ^2 (or σ_n^2) plays a significant role in CIE performance - the less bias the better. For instance, when m and b are fixed, the relative performance of the CIE's with respect to coverage is directly related to Bias($\hat{V}$). A secondary factor in CIE performance concerns the underlying marginal distribution of the X_i's. Further, with fixed m (and even more so with fixed n), coverage often deteriorates as b or α increase; this is partially due to the fact that, in these cases, $t_{d,1-\alpha/2}$ decreases.

Which CIE should one use in practice? If the sample size n is "large enough," we could probably argue successfully for CIE_N, CIE_O, or CIE_C with common d.o.f., $15 \leq d \leq 30$. In comparison to the other methods, CIE_N is probably the "safest" small sample method. There are several interesting research lines. We would like to see more emphasis on small sample results, including sequential procedures (which were not investigated here). Another question concerns the fact that for fixed d.o.f., CIE_N, CIE_O, and CIE_C achieve approximately valid coverage for about the same sample size. Further, a good batch size estimation procedure would be of tremendous import.

ACKNOWLEDGMENTS

We thank M.J. Rao and J. Chacko for some computer support. We are indebted to Bruce Schmeiser for many interesting discussions. This work was supported in part by the Air Force Systems Command, Rome Air Development Center, Griffiss Air Force Base, New York, and by Naval Postgraduate School, Monterey, California.

REFERENCES

Bratley, P., B.L. Fox, and L.E. Schrage. 1987. *A Guide to Simulation*, 2nd Ed., Springer-Verlag, New York.

Conway, R.W. 1963. "Some Tactical Problems in Digital Simulation." *Management Science*, 10, 47–61.

Crane, M.A. and D.L. Iglehart. 1975. "Simulating Stable Stochastic Systems: III. Regenerative Processes and Discrete-Event Simulations." *Operations Research*, 23, 33–45.

Crane, M.A. and A.J. Lemoine. 1977. *An Introduction to the Regenerative Method for Simulation Analysis*. Springer-Verlag, Berlin.

Damerdji, H. 1987. *Topics in Discrete-Event Stochastic Systems*. Ph.D. Dissertation, Dept. of Industrial Engineering, Univ. of Wisconsin, Madison, Wis.

Fishman, G.S. 1973. *Concepts and Methods in Discrete Event Digital Simulation*. John Wiley & Sons, New York.

Fishman, G.S. 1978. *Principles of Discrete Event Simulation*. Wiley, NY.

Glynn, P.W. 1982. "Coverage Error for Confidence Intervals Arising in Simulation Output Analysis." *Proc. 1982 Winter Simul. Conf.*, 369–375.

Glynn, P.W. and D.L. Iglehart. 1990. "Simulation Output Analysis Using Standardized Time Series." *Mathematics of Operations Research*, 15, 1–16.

Goldsman, D. 1984. *On Using Standardized Time Series to Analyze Stochastic Processes*. Ph.D. Dissertation, School of Operations Research and Industrial Engineering, Cornell Univ., Ithaca, NY.

Goldsman, D., K. Kang, and R.G. Sargent. 1986. "Large and Small Sample Comparisons of Various Variance Estimators." *Proc. 1986 Winter Simul. Conf.*, 278–284.

Goldsman, D. and M.S. Meketon. 1986. "A Comparison of Several Variance Estimators." Tech. Report #J–85–12, School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta.

Goldsman, D. and L. Schruben. 1984. "Asymptotic Properties of Some Confidence Interval Estimators for Simulation Output." *Management Science*, 30, 1217–1225.

Heidelberger, P. and P.D. Welch. 1981. "A Spectral Method for Confidence Interval Generation and Run Length Control in Simulations." *Comm. ACM*, 24, 233–245.

Heidelberger, P. and P.D. Welch. 1983. "Simulation Run Length Control in the Presence of an Initial Transient." *Operations Research*, 31, 1109–1144.

Kang, K. 1984. *Confidence Interval Estimation via Batch Means and Time Series Modeling*. Ph.D. Dissertation, School of Industrial Engineering, Purdue Univ., W. Lafayette, IN.

Kang, K. and D. Goldsman. 1990. "The Correlation between Mean and Variance Estimators in Computer Simulations." *IIE Transactions*, 22, 15–23.

Law, A.M. and W.D. Kelton. 1984. "Confidence Intervals for Steady-State Simulations, I: A Survey of Fixed Sample Size Procedures." *Operations Research*, 32, 1221–1239.

Lewis, P.A.W. 1980. "Simple Models for Positive-Valued and Discrete-Valued Time Series with ARMA Correlation Structure." *Multivariate Analysis–V*, P.R. Krishnaiah, ed., North Holland Publishing Company.

Meketon, M.S. 1980. *The Variance Time Curve: Theory, Estimation and Application*. Ph.D. Dissertation, School of Operations Research and Industrial Engineering, Cornell Univ., Ithaca, NY.

Meketon, M.S. and B.W. Schmeiser. 1984. "Overlapping Batch Means: Something for Nothing?" *Proc. 1984 Winter Simul. Conf.*, 227–230.

Moran, P.A.P. 1975. "The Estimation of Standard Errors in Monte Carlo Simulation Experiments." *Biometrika*, 62, 1–4.

Muirhead, R.J. 1982. *Aspects of Multivariate Statistical Theory*. Wiley, NY.

Priestley, M.B. 1982. *Spectral Analysis and Time Series*. Academic Press, NY.

Sargent, R.G., K. Kang, and D. Goldsman. 1987. "An Investigation of Small Sample Behavior of Confidence Interval Estimation Procedures." Working Paper #87-005, Dept. of Industrial Engineering and Operations Research, Syracuse Univ., Syracuse, NY 13244.

Schmeiser, B.W. 1982. "Batch Size Effects in the Analysis of Simulation Output." *Operations Research*, 30, 556–568.

Schmeiser, B.W. 1986. Personal Communication.

Schmeiser, B.W. and W.T. Song. 1989. "Optimal Mean-Squared-Error Batch Sizes." Tech. Report, School of Industrial Engineering, Purdue Univ., W. Lafayette, IN.

Schriber, T.J. and R.W. Andrews. 1984. "ARMA-Based Confidence Intervals for Simulation Output Analysis." *Amer. J. Math. and Mgmt. Sci.*, 4, 345–373.

Schruben, L. 1983. "Confidence Interval Estimation Using Standardized Time Series." *Operations Research*, 31, 1090–1108.

Song, W.T. and B.W. Schmeiser. 1989. "Estimating Variance of the Sample Mean: Quadratic Forms and Cross-Product Covariances." *Operations Research Letters*, 7, 259–266.

APPENDIX

Proof of (2)-(5): We will use the following facts:

$$\sum_{i=1}^{k} p^i = \frac{p(1-p^k)}{1-p} \quad \text{and} \quad \sum_{i=1}^{k} ip^i = \frac{p[1-(k+1)p^k+kp^{k+1}]}{(1-p)^2}. \tag{A-1}$$

For a covariance stationary process $\{X_i\}$ with $\gamma_k \equiv \mathrm{Cov}(X_i, X_{i+k})$,

$$\sigma_n^2 = n\cdot\mathrm{Var}(\bar{\bar{X}}_n) = \gamma_0 + \frac{2}{n}\sum_{i=1}^{n-1}(n-i)\gamma_i. \tag{A-2}$$

For the AR(1), $\gamma_k = \varphi^{|k|}$, and so

$$\sigma_n^2 = \frac{1+\varphi}{1-\varphi} - \frac{2\varphi(1-\varphi^n)}{n(1-\varphi)^2} \quad \text{and} \quad \sigma^2 = \lim_{n\to\infty}\sigma_n^2 = \frac{1+\varphi}{1-\varphi}. \tag{A-3}$$

Since

$$\frac{b-1}{m}E[\hat{V}_N] = \sum_{j=1}^{b} E[\bar{X}_{j,m}^2] - bE[\bar{\bar{X}}_n^2] = b\{E[\bar{\bar{X}}_m^2]-E[\bar{\bar{X}}_n^2]\} = b[\mathrm{Var}(\bar{\bar{X}}_m)-\mathrm{Var}(\bar{\bar{X}}_n)], \quad \text{(A-4)}$$

result (2) follows by (A-3) and simplification (see Moran 1975). //

We also have $\dfrac{(n-m+1)(n-m)}{nm} E[\hat{V}_0] = \sum_{i=1}^{n-m+1} E[(\bar{X}(i,m) - \bar{\bar{X}}_n)^2]$

$$= (n-m+1)[\mathrm{Var}(\bar{\bar{X}}_m) + \mathrm{Var}(\bar{\bar{X}}_n)] - 2\sum_{i=1}^{n-m+1} E[\bar{X}(i,m)\bar{\bar{X}}_n]. \quad \text{(A-5)}$$

Since $\mu = 0$, $mnE[\bar{X}(i,m)\bar{\bar{X}}_n] = \sum_{j=0}^{m-1} \sum_{k=1}^{n} E[X_{i+j}X_k] = \sum_{j=i}^{m+i-1} \sum_{k=1}^{n} \gamma_{j-k}$ (A-6)

$$= \sum_{j=i}^{m+i-1} \left\{\sum_{k=1}^{j} \varphi^{j-k} + \sum_{k=j+1}^{n} \varphi^{k-j}\right\} = \sum_{j=i}^{m+i-1} \frac{1 + \varphi - \varphi^j - \varphi^{n-j+1}}{1-\varphi}$$

$$= m\cdot\frac{1+\varphi}{1-\varphi} - (1-\varphi^m)(\varphi^i + \varphi^{n-m+2}\varphi^{-i})/(1-\varphi)^2. \quad \text{(A-7)}$$

We obtain (3) by substituting (A-3) and (A-7) into (A-5). //

(4) follows from (A.5–15) of Goldsman (1984); (5) follows from (2) and (4). //

<u>Proof of (6)-(9)</u>: The MA(1) has covariance function $\gamma_0 = 1 + \theta^2$, $\gamma_{\pm 1} = \theta$, and $\gamma_k = 0$, otherwise. By (A-2),

$$\sigma_n^2 = (1+\theta)^2 - 2\theta/n \quad \text{and} \quad \sigma^2 = \lim_{n\to\infty}\sigma_n^2 = (1+\theta)^2. \quad \text{(A-8)}$$

Result (6) then follows from (A-4) and (A-8). //

Note that

$$\sum_{k=1}^{n} \gamma_{j-k} = \begin{cases} (1+\theta)^2 & \text{if } 1 < j < n \\ (1+\theta)^2-\theta & \text{if } j = 1 \text{ or } n \end{cases}.$$

So by (A-6),

$$mnE[\bar{X}(i,m)\bar{\bar{X}}_n] = \begin{cases} m(1+\theta)^2 & \text{if } 1 < i < n-m+1 \\ m(1+\theta)^2-\theta & \text{if } j = 1 \text{ or } n-m+1 \end{cases} . \qquad \text{(A-9)}$$

We obtain (7) by substituting (A-8) and (A-9) into (A-5). //

(8) follows from (A.5-23) of Goldsman (1984); (9) follows from (6) and (8). //

Proof of independence in Examples 3, 4, and 5: Let $\underset{\sim}{X}' = (X_1,\ldots,X_n)$, where the X_i's are stationary normal with covariance matrix Σ. Suppose G is an $n \times n$ symmetric matrix and $\hat{V} = \underset{\sim}{X}'G\underset{\sim}{X}$. Problem 1.22 of Muirhead (1982) says that $\hat{V}$ and $\bar{\bar{X}}_n$ are independent iff $\underset{\sim}{1}'\Sigma G = \underset{\sim}{0}'$, where $\underset{\sim}{1}'$ [$\underset{\sim}{0}'$] is a $1 \times n$ vector of 1's [0's]. Song and Schmeiser (1989) note that $\hat{V}_N = \underset{\sim}{X}'G_N\underset{\sim}{X}$ and $\hat{V}_A = \underset{\sim}{X}'G_A\underset{\sim}{X}$, where

$$G_N = \begin{bmatrix} J & -J \\ -J & J \end{bmatrix} \text{ for } b = 2, \quad \text{and} \quad G_A = (h_i h_j) \text{ for } b = 1,$$

where J is an $n/2 \times n/2$ matrix consisting of 1's, and $h_i = (n+1)/2 - i$, $i = 1,\ldots,n$. Then G_N and G_A both meet Muirhead's condition for any Σ. //

Table 1 - Some Results for the AR(1) and EAR(1) Models with $\varphi = 0.9$. All entries for $E[\hat{V}]$ are exact. Results for CVG and CCVG are exact(*) or are based on 2000 independent simulation runs. For these models, $\sigma^2 = 19$.

	$E[\hat{V}]$ (or σ_n^2)		AR(1)				EAR(1)			
			b = 2		b = 16		b = 2		b = 16	
	b=2	b=16	CVG	CCVG	CVG	CCVG	CVG	CCVG	CVG	CCVG
m = 4										
NOBM	0.86	2.68	0.744*	0.900*	0.479	0.866	0.625	0.798	0.465	0.810
OBM	0.58	2.64	0.453	0.840	0.466	0.859	0.337	0.673	0.446	0.804
Area	0.31	0.31	0.418	0.907	0.183	0.888	0.296	0.701	0.164	0.799
Comb	0.49	1.46	0.432	0.878	0.365	0.862	0.321	0.683	0.337	0.806
(σ_n^2	6.19	16.19)								
m = 16										
NOBM	6.10	9.27	0.852*	0.900*	0.756	0.896	0.791	0.850	0.742	0.872
OBM	4.22	9.23	0.673	0.864	0.750	0.890	0.609	0.773	0.733	0.866
Area	2.85	2.85	0.684	0.896	0.491	0.891	0.615	0.800	0.480	0.866
Comb	3.94	5.96	0.689	0.884	0.651	0.893	0.625	0.795	0.646	0.868
(σ_n^2	13.57	18.30)								
m = 64										
NOBM	14.79	16.02	0.891*	0.900*	0.874	0.904	0.881	0.889	0.869	0.892
OBM	12.84	16.00	0.841	0.883	0.862	0.894	0.810	0.841	0.867	0.888
Area	11.29	11.29	0.849	0.897	0.801	0.894	0.822	0.866	0.803	0.884
Comb	12.45	13.58	0.852	0.895	0.840	0.898	0.819	0.858	0.841	0.888
(σ_n^2	17.59	18.82)								
m = 256										
NOBM	17.95	18.25	0.898*	0.900*	0.899	0.905	0.900	0.902	0.886	0.890
OBM	17.30	18.25	0.895	0.904	0.902	0.905	0.900	0.909	0.890	0.897
Area	16.90	16.90	0.895	0.902	0.886	0.903	0.888	0.895	0.878	0.896
Comb	17.25	17.56	0.890	0.901	0.897	0.908	0.896	0.902	0.880	0.892
(σ_n^2	18.65	18.96)								
m = 1024										
NOBM	18.74	18.81	0.900*	0.900*	0.897	0.899	0.883	0.883	0.909	0.910
OBM	18.56	18.81	0.905	0.907	0.897	0.898	0.915	0.918	0.902	0.903
Area	18.47	18.47	0.901	0.905	0.893	0.897	0.893	0.895	0.898	0.901
Comb	18.56	18.64	0.892	0.896	0.900	0.902	0.895	0.897	0.899	0.903
(σ_n^2	18.91	18.99)								

Table 2 - CVG Results for the AR(1) Process, b = 16, 1 - α = 0.90. All entries are based on at least 1000 independent simulation runs.

φ	-0.9	-0.5	0.0	0.5	0.9
m = 4					
NOBM	0.962	0.945	0.913	0.823	0.479
OBM	0.966	0.949	0.908	0.816	0.466
Area	1.000	0.974	0.898	0.677	0.183
Comb	0.998	0.971	0.900	0.755	0.365
m = 16					
NOBM	0.951	0.915	0.905	0.887	0.756
OBM	0.958	0.922	0.906	0.895	0.750
Area	0.988	0.942	0.908	0.860	0.491
Comb	0.981	0.926	0.910	0.880	0.651
m = 64					
NOBM	0.918	0.904	0.903	0.901	0.874
OBM	0.929	0.916	0.910	0.908	0.862
Area	0.943	0.901	0.885	0.885	0.801
Comb	0.931	0.900	0.896	0.894	0.840
m = 256					
NOBM	0.915	0.903	0.904	0.900	0.899
OBM	0.915	0.911	0.909	0.909	0.902
Area	0.931	0.914	0.918	0.917	0.886
Comb	0.921	0.918	0.916	0.915	0.897
m = 1024					
NOBM	0.906	0.901	0.899	0.899	0.897
OBM	0.903	0.899	0.897	0.897	0.897
Area	0.907	0.904	0.906	0.904	0.893
Comb	0.909	0.906	0.902	0.900	0.900

Table 3 - CVG Results for Variance Estimators with Common d.o.f.
All entries are based on 2000 independent simulation runs.

n =	32	64	128	256	512	1024	2048	4096	8192
AR(1), φ = 0.9, d = 3									
NOBM	0.688	0.787	0.848	0.880	0.889	0.899	0.896	0.902	0.905
OBM	0.692	0.792	0.834	0.874	0.896	0.910	0.903	0.923	0.920
Area	0.505	0.674	0.795	0.858	0.880	0.900	0.895	0.889	0.912
Comb	0.701	0.799	0.847	0.886	0.883	0.903	0.895	0.905	0.908
EAR(1), φ = 0.9, d = 3									
NOBM	0.622	0.736	0.826	0.862	0.887	0.900	0.896	0.903	0.896
OBM	0.631	0.725	0.820	0.864	0.887	0.906	0.917	0.912	0.916
Area	0.426	0.630	0.771	0.844	0.868	0.886	0.893	0.893	0.893
Comb	0.636	0.740	0.822	0.863	0.881	0.898	0.894	0.902	0.896
M/M/1, ρ = 0.8, d = 3									
NOBM	0.437	0.541	0.632	0.707	0.768	0.804	0.836	0.862	0.877
OBM	0.432	0.540	0.627	0.704	0.752	0.799	0.842	0.867	0.886
Area	0.290	0.420	0.533	0.624	0.708	0.769	0.817	0.835	0.855
Comb	0.442	0.548	0.634	0.701	0.758	0.806	0.838	0.860	0.871
MA(1), θ = -0.9, d = 3									
NOBM	0.985	0.981	0.986	0.975	0.953	0.932	0.931	0.920	0.911
OBM	1.000	1.000	1.000	1.000	0.998	0.986	0.971	0.959	0.944
Area	0.994	0.993	0.989	0.980	0.967	0.955	0.947	0.933	0.918
Comb	0.994	0.992	0.989	0.976	0.963	0.932	0.933	0.919	0.908
AR(1), φ = 0.9, d = 15									
NOBM	0.381	0.485	0.642	0.749	0.830	0.877	0.890	0.900	0.908
OBM	0.377	0.515	0.684	0.793	0.847	0.880	0.893	0.897	0.910
Area	0.126	0.189	0.318	0.513	0.695	0.809	0.871	0.891	0.903
Comb	0.385	0.491	0.649	0.755	0.834	0.876	0.894	0.903	0.907
EAR(1), φ = 0.9, d = 15									
NOBM	0.324	0.470	0.607	0.749	0.811	0.876	0.903	0.897	0.897
OBM	0.319	0.505	0.660	0.786	0.833	0.877	0.899	0.891	0.902
Area	0.076	0.173	0.303	0.496	0.670	0.802	0.868	0.882	0.887
Comb	0.326	0.479	0.615	0.754	0.814	0.879	0.893	0.894	0.898
M/M/1, ρ = 0.8, d = 15									
NOBM	0.227	0.318	0.386	0.490	0.618	0.701	0.773	0.819	0.857
OBM	0.226	0.342	0.425	0.525	0.653	0.725	0.779	0.820	0.858
Area	0.055	0.091	0.144	0.228	0.369	0.520	0.647	0.757	0.806
Comb	0.229	0.319	0.392	0.498	0.624	0.706	0.778	0.823	0.850
MA(1), θ = -0.9, d = 15									
NOBM	1.000	1.000	1.000	1.000	0.999	0.993	0.983	0.961	0.948
OBM	1.000	1.000	1.000	1.000	1.000	0.997	0.983	0.962	0.952
Area	1.000	1.000	1.000	1.000	1.000	1.000	0.997	0.990	0.978
Comb	1.000	1.000	1.000	1.000	0.998	0.992	0.980	0.963	0.950

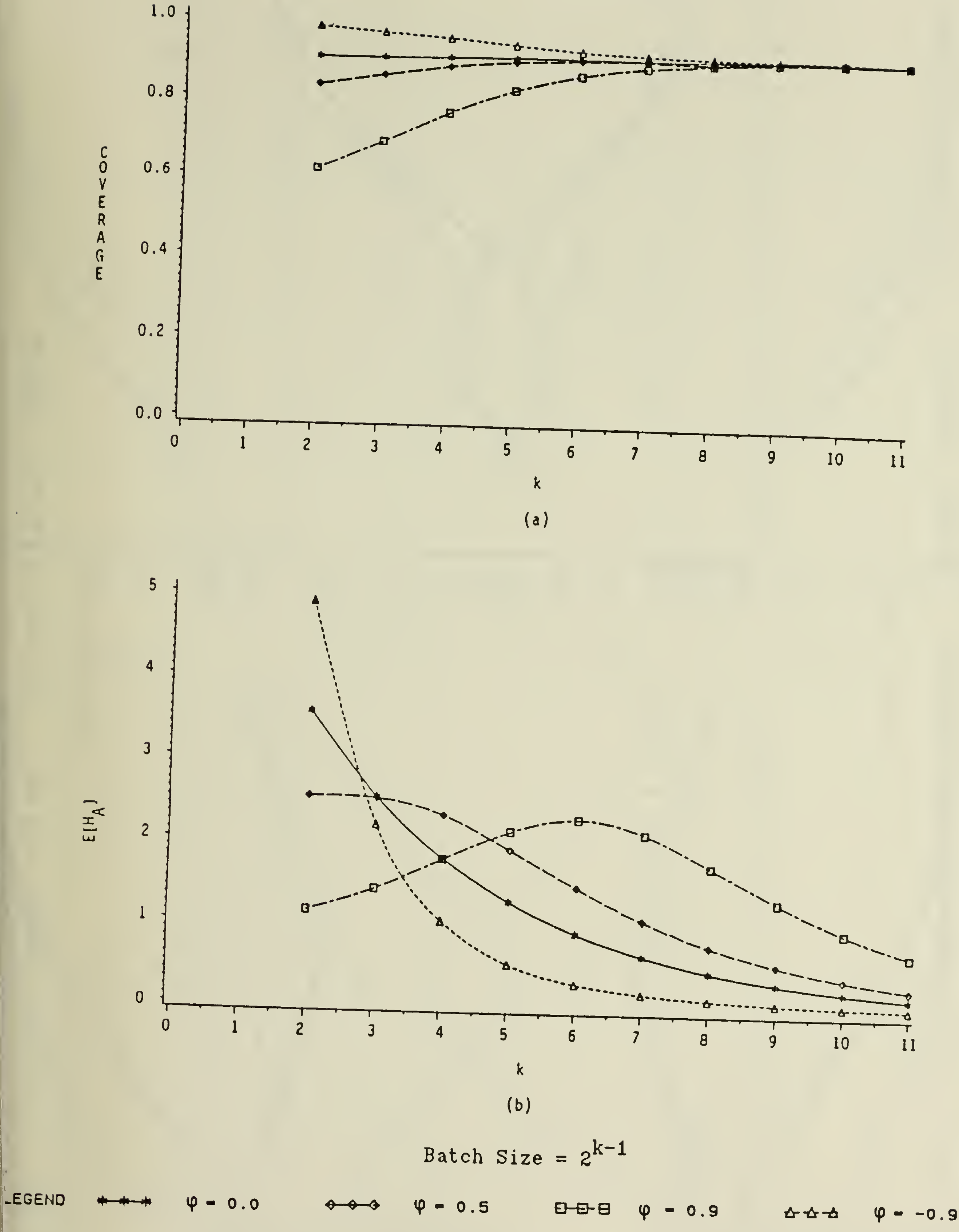
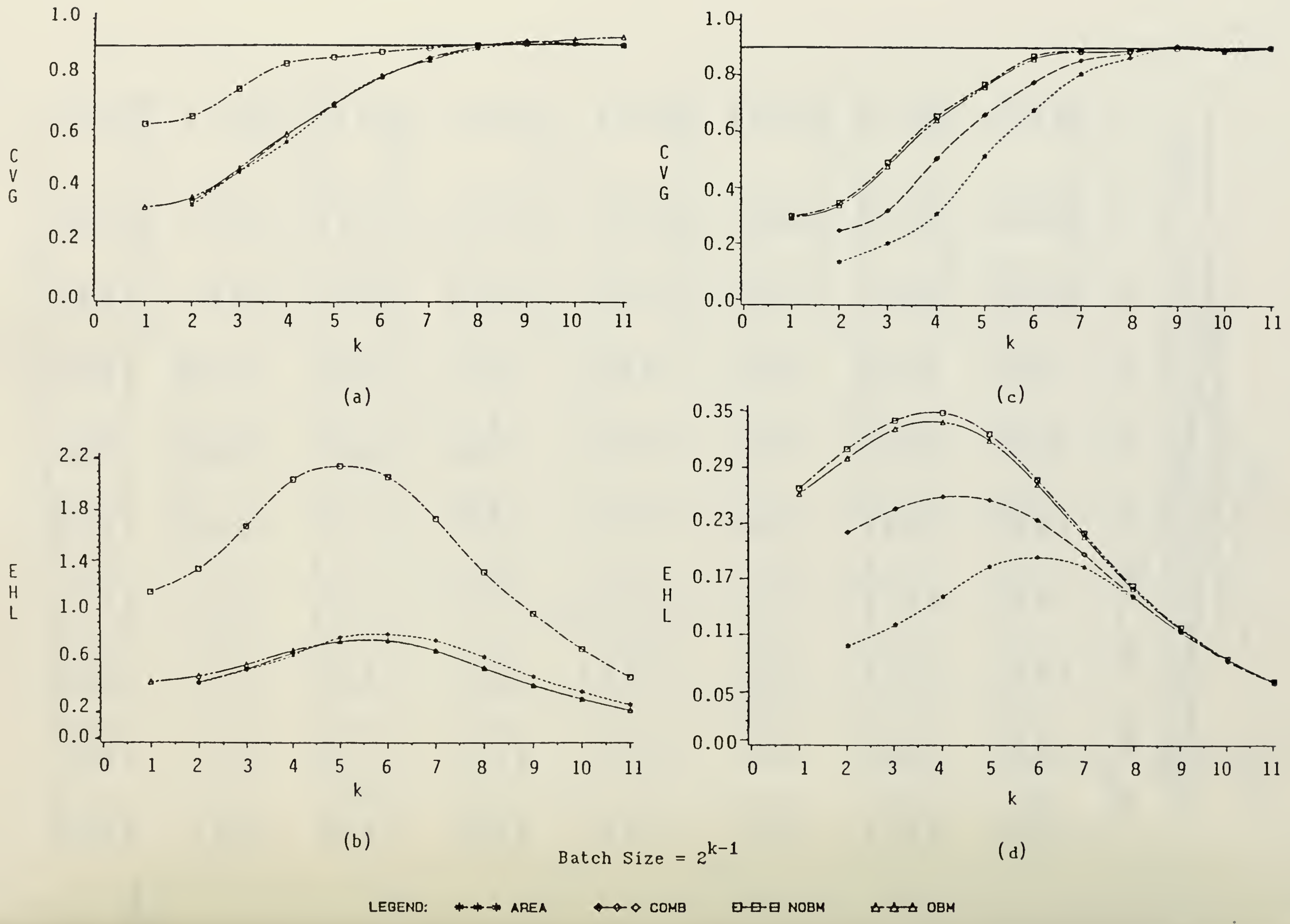


Figure 1: AR(1) Process Performance Measures for AREA CIE, b = 1


CVG
k
(a)
CVG
k
(c)
EHL
k
(b)
EHL
k
(d)
Batch Size = 2^{k-1}
LEGEND: AREA COMB NOBM OBM

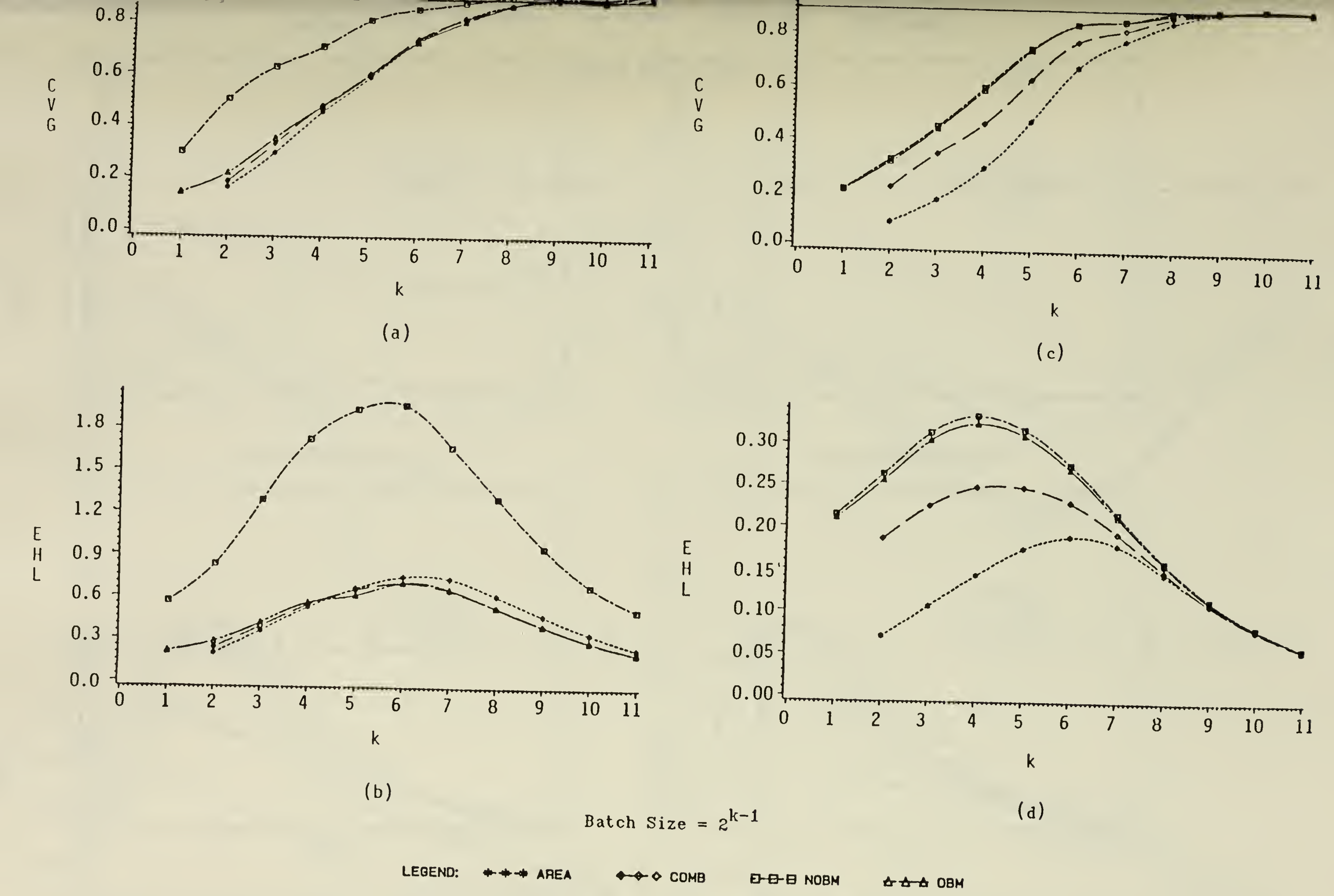
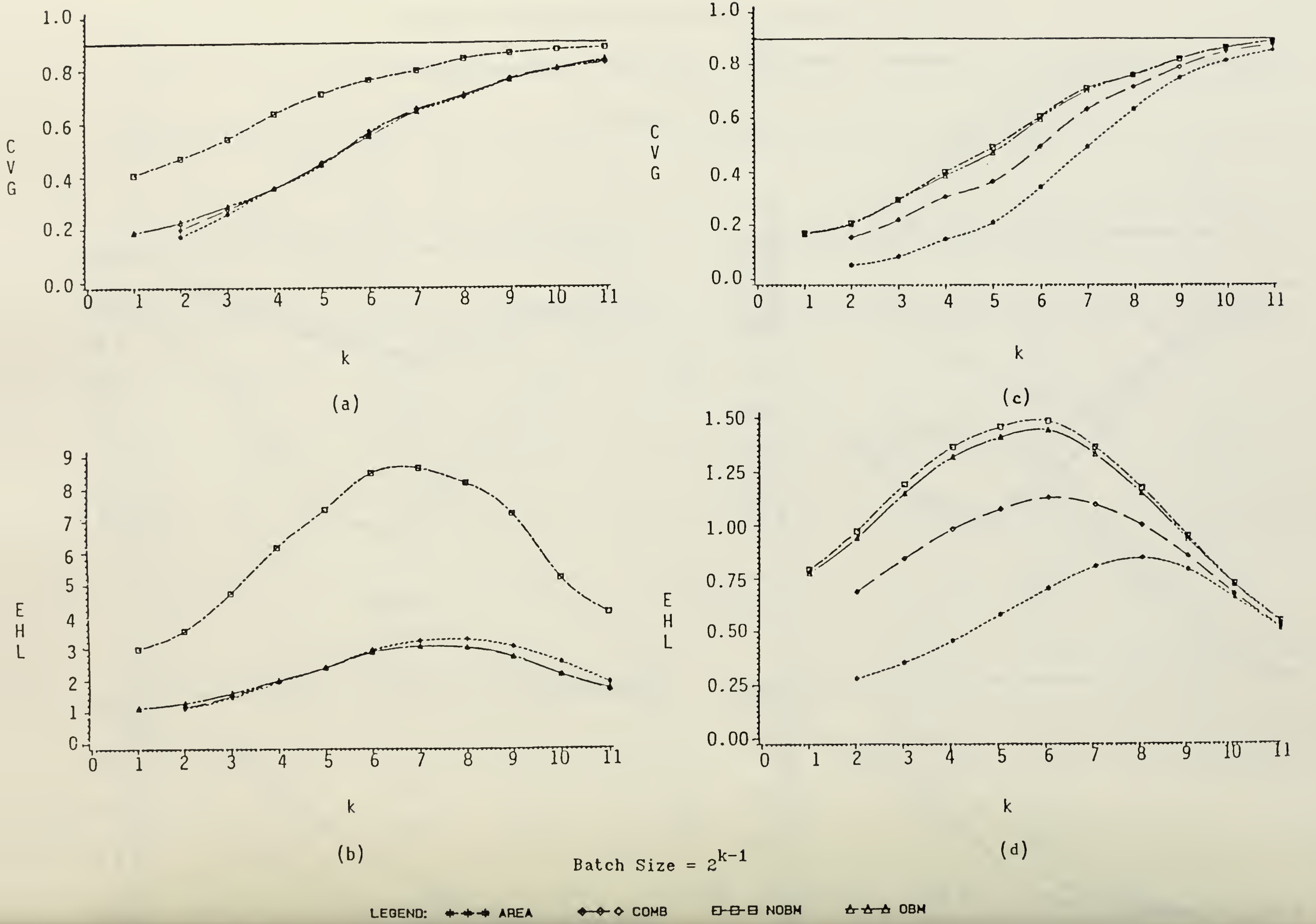


Figure 3: EAR(1) Process Performance Measures, $\varphi = 0.9$


C V G
k
(a)
E H L
k
(b)
C V G
k
(c)
E H L
k
(d)
Batch Size = 2^{k-1}
LEGEND: AREA COMB NOBM OBM

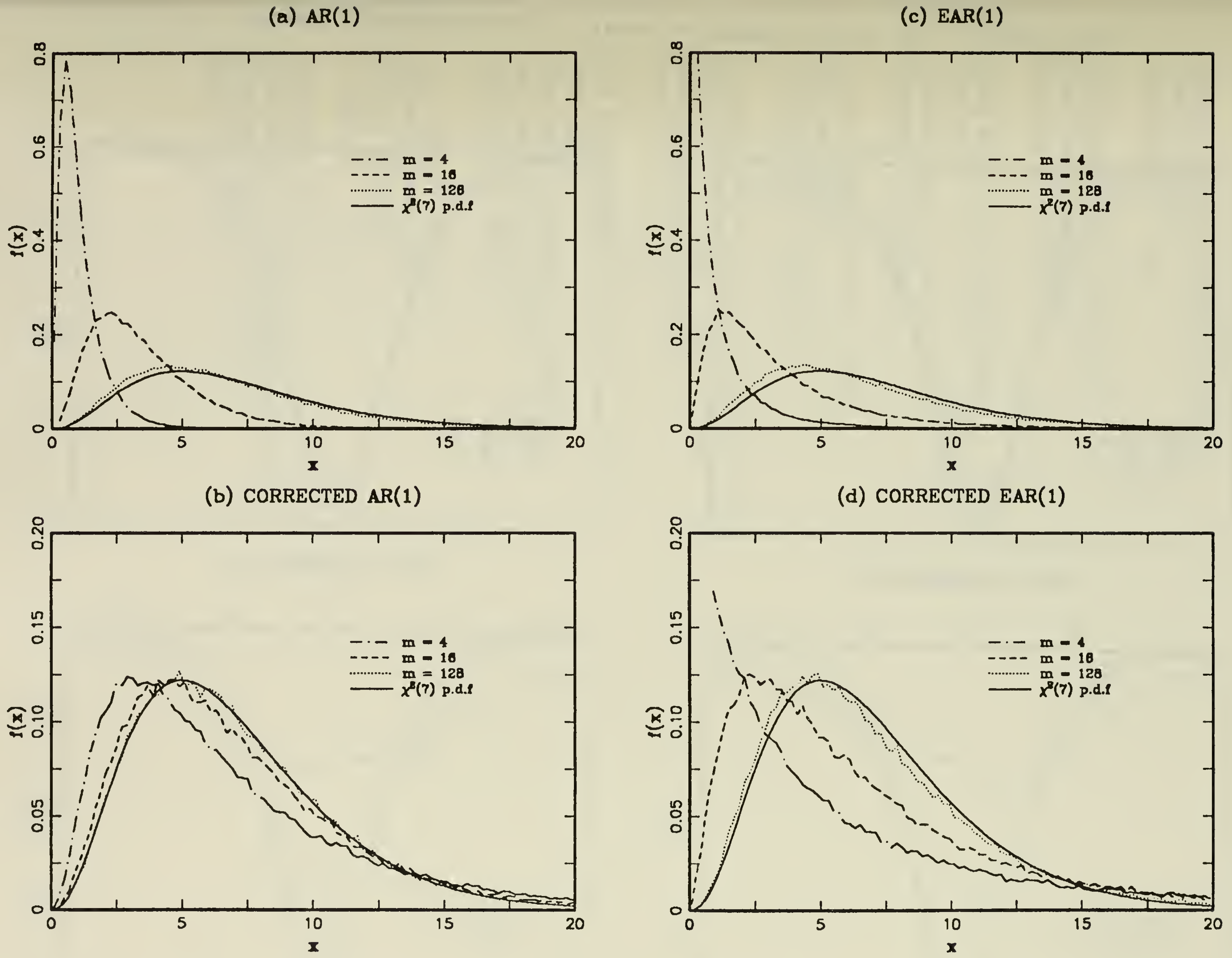


Figure 5: Empirical p.d.f.'s (based on 100000 runs) of $\hat{V}_N/\sigma_n^2$ and $\tilde{V}_N/\sigma_n^2$ for the AR(1) and EAR(1) with $\varphi = 0.9$ and $b = 8$

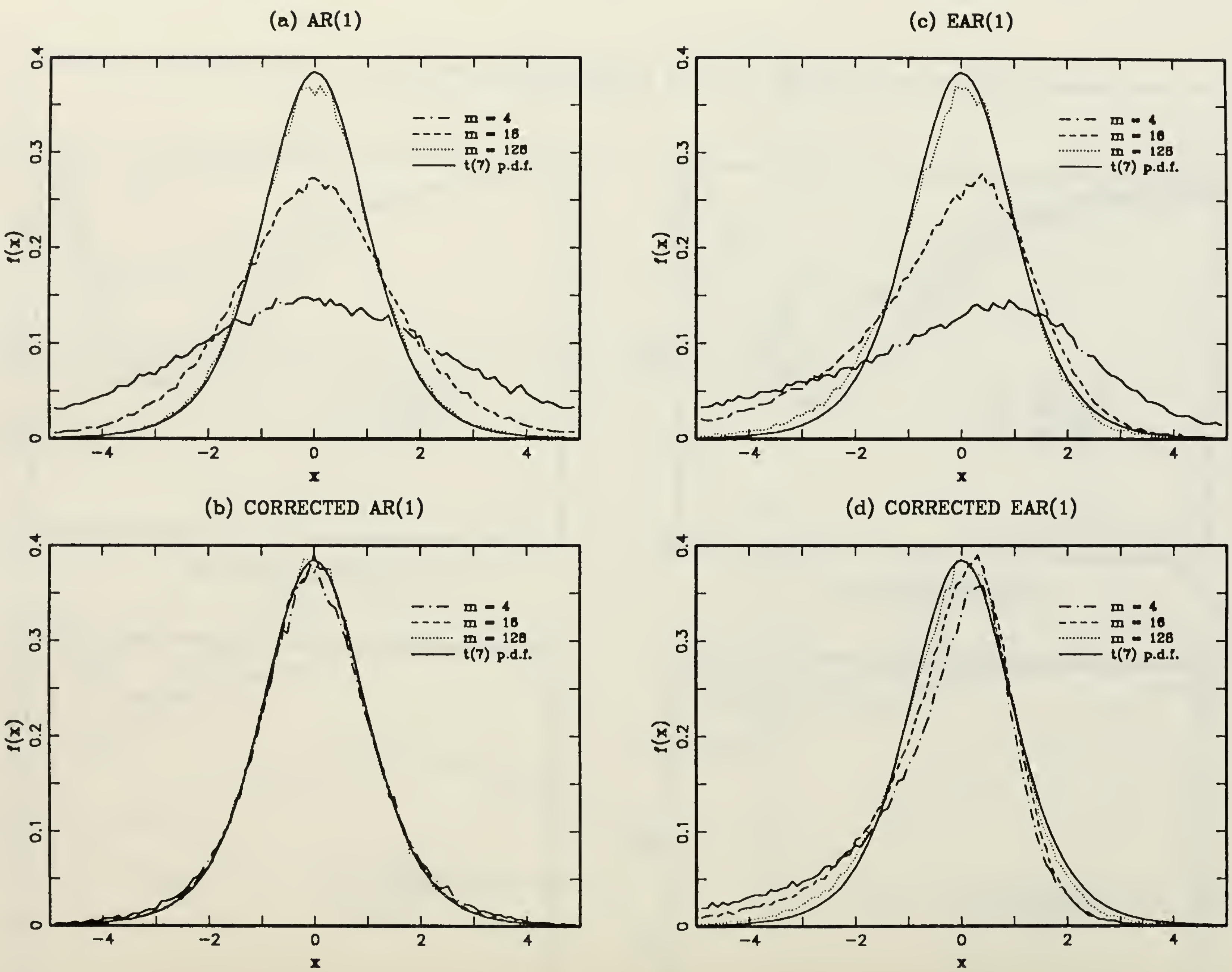


Figure 6: Empirical p.d.f.'s (based on 100000 runs) of $(\bar{\bar{X}}_n - \mu)/(\hat{V}_N/n)^{1/2}$ and $(\bar{\bar{X}}_n - \mu)/(\tilde{V}_N/n)^{1/2}$ for the AR(1) and EAR(1) with $\varphi = 0.9$ and $b = 8$.

Distribution List

Agency	No. of copies
Defense Technical Information Center Cameron Station Alexandria, VA 22314	2
Dudley Knox Library, Code 0142 Naval Postgraduate School Monterey, CA 93943	2
Office of Research Administration Code 08 Naval Postgraduate School Monterey, CA 93943	1
Library, Center for Naval Analyses 4401 Ford Avenue Alexandria, VA 22302-0268	1
Department of Administrative Sciences Library Code AS Naval Postgraduate School Monterey, CA 93943	1
Prof. Keebom Kang Code AS/Kk Administrative Sciences Department Naval Postgraduate School Monterey, Ca. 93943	10
Prof. Robert G. Sargent Simulation Research Group Syracuse University Syracuse, NY 13244	20
Prof. David Goldsman School of Industrial & Systems Engineering Georgia Institute of Technology Atlanta, GA 30332	10