SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER: 4	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle): Simple and Efficient Estimation of Parameters of Non-Gaussian Autoregressive Processes		5. TYPE OF REPORT & PERIOD COVERED: August 1984 to July 1986 6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s): Steven Kay and Debasis Sengupta		8. CONTRACT OR GRANT NUMBER(s): N00014-84-K-0527
9. PERFORMING ORGANIZATION NAME AND ADDRESS: Electrical Engineering Department, University of Rhode Island, Kingston, Rhode Island 02881		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS: Office of Naval Research, Department of the Navy, Arlington, Virginia 22217		12. REPORT DATE: August 1986 13. NUMBER OF PAGES: 42
14. MONITORING AGENCY NAME & ADDRESS (If different from Controlling Office)		15. SECURITY CLASS. (of this report): Unclassified 15a. DECLASSIFICATION DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release. Distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Parameter Estimation
Autoregressive processes
Non-Gaussian processes
Maximum Likelihood Estimator
Weighted Least Squares
Efficiency Robustness

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

A new technique for the estimation of autoregressive filter parameters of a non-Gaussian autoregressive process is proposed. The probability density function of the driving noise is assumed to be known. The new technique is a two-stage procedure motivated by maximum likelihood estimation. It is computationally much simpler than the maximum likelihood estimator and does not suffer from convergence problems. Computer simulations indicate that unlike the least squares or linear prediction estimators, the proposed estimator is nearly efficient, even for moderately sized data records. By a

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

S. N 0102-LF-014-6601

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

Report Number 4

Simple and Efficient Estimation of Parameters of Non-Gaussian Autoregressive Processes

STEVEN KAY AND DEBASIS SENGUPTA

Department of Electrical Engineering
University of Rhode Island
Kingston, Rhode Island 02881

August 1986

Prepared for

OFFICE OF NAVAL RESEARCH
Statistics and Probability Branch
Arlington, Virginia 22217
under Contract N00014-84-K-0527
S.M. Kay, Principal Investigator

Abstract

A new technique for the estimation of autoregressive filter parameters of a non-Gaussian autoregressive process is proposed. The probability density function of the driving noise is assumed to be known. The new technique is a two-stage procedure motivated by maximum likelihood estimation. It is computationally much simpler than the maximum likelihood estimator and does not suffer from convergence problems. Computer simulations indicate that unlike the least squares or linear prediction estimators, the proposed estimator is nearly efficient, even for moderately sized data records. By a slight modification the proposed estimator can also be used in the case when the parameters of the driving noise probability density function are not known.

I. Introduction

Estimation of the parameters of autoregressive (AR) processes has been widely addressed [Box and Jenkins 1970], [Kay 1986]. These processes are modeled by an all-pole filter excited by a white Gaussian process, also referred to as the *driving noise*. The class of AR processes driven by white *non*-Gaussian noise has not received much attention, although they are capable of representing a wide range of physical processes [Sengupta and Kay 1986]. Previous research has shown that it may be possible to estimate some of the parameters characterizing a non-Gaussian AR process more precisely than those of a Gaussian AR process with the same power spectral density (PSD). Specifically, the Cramer-Rao bound (CR bound) for the variances of the estimators for the AR filter parameters [Martin 1982] is lower in the non-Gaussian case than in the Gaussian case [Pakula 1986], [Sengupta 1986]. Yet utilization of this theoretical result has been limited. The method of maximum likelihood, which is the most widely considered approach for AR parameter estimation, is usually associated with computational complexity and convergence problems. In the case of non-Gaussian PDF's maximization of the likelihood function leads to a set of highly non-linear equations [Sengupta and Kay 1986] in contrast to the Gaussian case where the equations become linear after a few simplifying assumptions. It is the solution of these non-linear equations which results in the computational complexity of the maximum likelihood estimator (MLE). Iterative techniques which are often used to solve these equations suffer from convergence problems for short data records.

Several non-Gaussian PDF's have been proposed to model the driving noise. Zero-mean symmetric PDF's having tails heavier than the Gaussian tail are of particular interest because they may model a nominally Gaussian background with occasional impulses. Such noise processes are encountered in low-frequency atmospheric communications [Bernstein 1974], sonar and radar problems and so on. Examples of heavy-tailed PDF's are the mixed-Gaussian PDF, the Johnson family and Middleton's class A and

B PDF families. All of them lead to a complicated likelihood function which is difficult to maximize.

This paper suggests a method to estimate the AR filter parameters of a non-Gaussian process in a computationally simple way. Essentially it is a two-stage procedure based on an approximation of the MLE. The resulting estimator is asymptotically efficient in the sense that its variance approaches the CR bound for large data records. The mixed-Gaussian PDF is used to illustrate the approach and to demonstrate some of the finer aspects.

The paper is organized as follows. Section II gives an interpretation of the MLE which forms the basis for the development of the new estimator. Several special cases are discussed to illustrate the central argument. Section III suggests an approximation to the MLE based on this interpretation. Section IV actually implements such an estimator for the case of a mixed-Gaussian distribution. Section V discusses the case when some of the parameters of the noise PDF are unknown while section VI reports the results of computer simulations. Section VII summarizes the main results.

II. An Interpretation of the MLE

Consider N observations of an AR(p) process

$$x_n = -\sum_{j=1}^{p} a_j x_{n-j} + u_n, \qquad n = 1, 2, \cdots N \tag{1}$$

where the driving noise u_n has the PDF $f(u_n; \Theta)$ dependent on the parameter vector Θ. f is assumed to be an even function and hence zero mean. It is also assumed that f has tails heavier than a Gaussian PDF g having equal variance, *i.e.*, there is a number U such that

$$f(u_n) > g(u_n) \qquad \text{for } |u_n| > U \tag{2}$$

subject to the constraint

$$\int_{-\infty}^{+\infty} u_n^2 \, f(u_n) du_n = \int_{-\infty}^{+\infty} u_n^2 \, g(u_n) du_n$$

The log likelihood function for the AR filter parameters is given by the joint PDF of $\{x_1, x_2, \cdots, x_N\}$ when the x_i's are replaced by their observed values. This is difficult to evaluate. It is a common practice to replace the exact likelihood function by the conditional liklihood of $\{x_{p+1}, x_{p+2}, \cdots, x_N\}$ given $\{x_1, x_2, \cdots, x_p\}$ for the purpose of maximization over the parameters. It can be easily shown [Box and Jenkins 1970] that the conditional log likelihood function is given by

$$\ln \mathbf{f} = \sum_{n=p+1}^{N} \ln f(u_n; \Theta)\Big|_{u_n = \sum_{i=0}^{p} a_i x_{n-i}} \tag{3}$$

where $a_0 = 1$. Differentiating with respect to a_j

$$\begin{aligned}
\frac{\partial}{\partial a_j} \ln \mathbf{f} &= \sum_{n=p+1}^{N} \frac{\partial}{\partial u_n} \ln f(u_n; \Theta) \frac{\partial u_n}{\partial a_j}\Big|_{u_n = \sum_{i=0}^{p} a_i x_{n-i}} \\
&= \sum_{n=p+1}^{N} x_{n-j} \frac{f'(u_n; \Theta)}{f(u_n; \Theta)}\Big|_{u_n = \sum_{i=0}^{p} a_i x_{n-i}} \\
&= \sum_{n=p+1}^{N} x_{n-j} u_n \frac{f'(u_n; \Theta)}{u_n f(u_n; \Theta)}\Big|_{u_n = \sum_{i=0}^{p} a_i x_{n-i}} \\
&= -\sum_{n=p+1}^{N} x_{n-j} u_n \Gamma(u_n)\Big|_{u_n = \sum_{i=0}^{p} a_i x_{n-i}}
\end{aligned} \tag{4}$$

where

$$\Gamma(u_n) = -\frac{f'(u_n; \Theta)}{u_n f(u_n; \Theta)} \tag{5}$$

The MLE of $\mathbf{a} = [a_1 \ a_2 \ \cdots \ a_p]$ is found by solving

$$\sum_{n=p+1}^{N} x_{n-j} u_n \Gamma(u_n)\Big|_{u_n = \sum_{i=0}^{p} a_i x_{n-i}} = 0, \qquad j = 1, 2, \cdots p \tag{6}$$

provided Θ is either known or replaced by its MLE $\hat{\Theta}$ in order to calculate $\Gamma(u_n)$ from (5). From this point onwards Θ will be assumed to be known. It will be shown in section V that for some PDF's the method to be described can be implemented with

Θ replaced by a reasonable estimate. Note that since f is assumed to be a symmetric PDF, f' is an odd function of u_n. Therefore f'/u_n is even in u_n and $\Gamma(u_n)$ given by (5) is an even function. $\Gamma(u_n)$ is assumed to exist over the domain of $f(u_n;\Theta)$. (6) can also be written as

$$\sum_{i=0}^{p} a_i \sum_{n=p+1}^{N} x_{n-i}x_{n-j}\Gamma(u_n)\Big|_{u_n=\sum_{i=0}^{p} a_i x_{n-i}} = 0, \qquad j = 1,2,\cdots p \tag{7}$$

For the special case of a Gaussian PDF with variance σ^2, $f'/f = -u_n/\sigma^2$. Therefore $\Gamma(u_n) = 1/\sigma^2$ and (7) reduces to

$$\sum_{i=0}^{p} a_i \sum_{n=p+1}^{N} x_{n-i}x_{n-j} = 0, \qquad j = 1,2,\cdots p \tag{8}$$

which can be recognized as the *covariance method* of linear prediction, known to be approximately the MLE in the Gaussian case. It is the solution of a linear least squares (LS) problem [Box and Jenkins 1970], [Kay 1986]

$$\min_{\mathbf{a}} \sum_{n=p+1}^{N} \left(x_n + \sum_{j=1}^{p} a_j x_{n-j} \right)^2 \tag{9}$$

(7) resembles the solution of a LS problem except for the weighting factors $\Gamma(u_n)$. Note that the inherent dependence of u_n (and hence $\Gamma(u_n)$) on a, the AR filter parameters, makes it a non-linear problem. If, however, the argument of Γ is calculated using a *fixed* (and hopefully an approximate) value of a, (7) becomes the solution of the following *weighted* LS problem

$$\min_{\mathbf{a}} \sum_{n=p+1}^{N} \left(x_n + \sum_{j=1}^{p} a_j x_{n-j} \right)^2 \Gamma(\hat{u}_n) \tag{10}$$

$\hat{u}_n$, a quantity expected to be close to u_n is defined by (1)

$$\hat{u}_n = \sum_{j=0}^{p} \hat{a}_j x_{n-j}, \qquad n = p+1, p+2, \cdots N \tag{11}$$

where some fixed approximate values of the AR filter parameters are used. $\hat{a}_0$ is defined to be unity. (10) is minimized by (7) with $\Gamma(u_n)$ replaced by $\Gamma(\hat{u}_n)$ which reduces to a set of linear equations whose unique solution can be expected to be close to the MLE of a. The resulting estimator should be much better than the unweighted LS estimator (resulting from (9)) because it retains the general shape of $\Gamma(u_n)$ by approximating it with $\Gamma(\hat{u}_n)$. The sequence $\{\hat{u}_n \mid p+1 \leq n \leq N\}$ can be generated by passing $\{x_{p+1}, x_{p+2}, \cdots, x_N\}$ through a moving average (MA) filter as per (11) with coefficients obtained from a preliminary stage of least squares estimation (*e.g.*, by covariance method). In other words, $\hat{u}_n$ becomes an estimate of the nth sample of the driving noise based on an LS estimate of the filter parameters such as the covariance method. This approach leads to an approximate MLE which is described in the next section.

It is of interest to know how the terms of (10) are actually weighted. Three symmetric PDF's which are commonly used to model heavy-tailed non-Gaussian processes [Czarnecki and Thomas 1982], [Middleton 1977], [Johnson and Kotz 1971] are now considered. The plots of the weighting function $\Gamma(u)$ for these PDF's provide insight into the structure of the MLE.

Mixed-Gaussian Model: The mixed-Gaussian PDF has received considerable attention in situations where the underlying random process is characterized by the presence of occasional impulses in an otherwise Gaussian process. The PDF is given by

$$f(u) = (1-\epsilon)E_B(u) + \epsilon E_I(u), \qquad 0 < \epsilon < 1 \tag{12}$$

where E_B and E_I are Gaussian PDF's with parameters $[\mu_B, \sigma_B^2]$ and $[\mu_I, \sigma_I^2]$, respectively. Assuming $\sigma_B^2 << \sigma_I^2$, the fraction ϵ can be thought of as the degree of contamination of the low-variance Gaussian process with PDF E_B by the high-variance component with PDF E_I. Only the zero-mean case ($\mu_B = 0, \mu_I = 0$) will be considered

here. (12) is explicitly written as

$$f(u) = \frac{1-\epsilon}{\sqrt{2\pi\sigma_B^2}} e^{-\frac{u^2}{2\sigma_B^2}} + \frac{\epsilon}{\sqrt{2\pi\sigma_I^2}} e^{-\frac{u^2}{2\sigma_I^2}} \tag{13}$$

For this PDF $\Gamma(u)$ can be shown to be

$$\begin{aligned}\Gamma(u) &= \frac{1}{\sigma_B^2} \frac{\frac{1-\epsilon}{\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{u}{\sigma_B}\right)^2} + \frac{\epsilon}{\rho\sqrt{2\pi\rho}} e^{-\frac{1}{2\rho}\left(\frac{u}{2\sigma_B}\right)^2}}{\frac{1-\epsilon}{\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{u}{\sigma_B}\right)^2} + \frac{\epsilon}{\sqrt{2\pi\rho}} e^{-\frac{1}{2\rho}\left(\frac{u}{2\sigma_B}\right)^2}} \\ &= \frac{1}{\sigma_B^2} \frac{\frac{1-\epsilon}{\sqrt{2\pi}} e^{-\frac{\dot{u}^2}{2}} + \frac{\epsilon}{\rho\sqrt{2\pi\rho}} e^{-\frac{\dot{u}^2}{2\rho}}}{\frac{1-\epsilon}{\sqrt{2\pi}} e^{-\frac{\dot{u}^2}{2}} + \frac{\epsilon}{\sqrt{2\pi\rho}} e^{-\frac{\dot{u}^2}{2\rho}}}\end{aligned} \tag{14}$$

where $\rho = \sigma_I^2/\sigma_B^2$ and $\dot{u} = u/\sigma_B$. Figure 1(a) plots $\Gamma(u)/\Gamma(0)$ vs. $\dot{u}$ (= u normalized by σ_B) for $\rho = 100$. Curves for different values of ϵ are overlayed. Figure 1(b) plots $\Gamma(u)/\Gamma(0)$ vs. $\dot{u}$ for $\epsilon = 0.1$ and different values of ρ. The curves show that $\Gamma(u)$ acts as a limiter. The squared errors in (10) with large values of $\hat{u}_n$ are suppressed. This makes intuitive sense because large values of u_n (*spikes* at the input of the AR filter) would otherwise dominate the sum of the squares and consequently the information contained in the rest of the terms will be lost. Quantitatively, from (14) with $\rho >> 1$, $\Gamma(0) \approx 1/\sigma_B^2$ and $\Gamma(u) \to 1/\sigma_I^2$ as $u \to \infty$, *i.e.*, very small and very large terms in (10) are scaled in the inverse ratio of background and interference noise powers, respectively. This is in accordance with analogous results in optimal weighted least squares theory [Sorenson 1980].

Middleton's Class A Model: Another physically motivated model to represent nominally Gaussian noise with an impulsive component is Middleton's class A PDF given by the infinite sum

$$f(u) = e^{-A} \sum_{m=0}^{\infty} \frac{A^m E_m(u)}{m!} \tag{15}$$

where $0 < A < 1$ and $E_m(u)$ is a zero-mean Gaussian PDF with variance σ_m^2 given by

$$\sigma_m^2 = \sigma^2 \frac{m/A + B}{1 + B} \tag{16}$$

with B a constant. The Gaussian component corresponding to $m = 0$ has the least variance and can be thought of as the *background* process. The weights of the higher order terms can be controlled by the constant A. A small value of A will diminish the contribution of the *contaminating* high-order terms. The constant B can be used to adjust the variance of these components. The overall variance of the Middleton's class A PDF is σ^2. $\Gamma(u)$ can be written in this case as

$$\Gamma(u) = \frac{\sum_{m=0}^{\infty} \frac{A^m}{\sigma_m^2 m!} E_m(u)}{\sum_{m=0}^{\infty} \frac{A^m}{m!} E_m(u)} = \frac{\sum_{m=0}^{\infty} \frac{A^m}{\sigma_m^3 m!} e^{-\frac{1}{2}\left(\frac{u}{\sigma_m}\right)^2}}{\sum_{m=0}^{\infty} \frac{A^m}{\sigma_m m!} e^{-\frac{1}{2}\left(\frac{u}{\sigma_m}\right)^2}} \tag{17}$$

Figures 2(a) and 2(b) plot $\Gamma(u)/\Gamma(0)$ vs. u for this PDF for different values of A and B. σ^2 is assumed to be unity. These curves also are observed to be similar to a limiter curve. A larger value of A implies an increased presence of high-variance components. Therefore a smaller threshold is necessary above which the squared terms of (10) need to be down-weighted in order to preserve the information in the remaining terms. This is reflected in Figure 2(a) which clearly exhibits a smaller threshold for a higher A. Also, a larger value of B implies less difference in the variances of the Gaussian terms corresponding to $m = 0$ and $m > 0$, *i.e.*, a smaller deviation from Gaussianity. A smaller value of B indicates more non-Gaussianity and hence a smaller threshold is necessary, which is confirmed by Figure 2(b).

Johnson Family: The Johnson family of PDF's is one of the heavy-tailed families obtained by applying a transformation to a Gaussian random variable. If v is a Gaussian random variable with mean zero and variance one, then the transformed random variable

$$u = t \sinh\left(\frac{v}{\delta}\right)$$

has a PDF belonging to the Johnson family given by

$$f(u) = \frac{\delta}{t\sqrt{2\pi}} \left[\frac{u^2}{t^2} + 1\right]^{-\frac{1}{2}} e^{-\frac{1}{2}\left(\delta \sinh^{-1}\left(\frac{u}{t}\right)\right)^2} \tag{18}$$

t is chosen to be

$$t = \left[\frac{2\sigma^2}{e^{\left(\frac{2}{\delta^2}\right)} - 1}\right]^{\frac{1}{2}}$$

so that the density has a variance σ^2. The parameter t can be used to control the heaviness of tail. A smaller value of t implies a heavier tail. The PDF approaches a Gaussian one as $\delta \to \infty$. $\Gamma(u)$ can be written in this case as

$$\Gamma(u) = \frac{1}{t^2}\left[1 + \frac{u^2}{t^2}\right]^{-1} + \left[1 + \frac{u^2}{t^2}\right]^{-\frac{1}{2}} \frac{\delta^2}{ut} \sinh^{-1}\left(\frac{u}{t}\right) \tag{19}$$

Figure 3 plots $\Gamma(u)/\Gamma(u)$ vs. u/σ for different values of t. A larger value of t indicates less deviation from Gaussianity confirmed by a gradual decrease of the curve from the value at $u = 0$. Smaller values of t correspond to a sharper transition and a smaller threshold.

All these illustrations show that the MLE given as a solution of (7) actually *down*weights the larger squared terms and may therefore be well approximated by an appropriate weighted LS estimator. The following section elaborates on this point. It should also be noted that $\Gamma(u)$ is positive in all the above cases. This is true for any symmetric PDF which is a monotonically decreasing function of u for positive values of u, as may be verified from (5). Most of the common PDF's have this property.

III. An approximation to the MLE

It was suggested in the previous section that the problem of solving the set of highly non-linear equations (7) can be replaced by solving a set of *linear* equations if the weighting function $\Gamma(u_n)$ is known or can be estimated for each n. Specifically, this suggests a two-stage procedure. The first stage involves computation of the unweighted LS estimates of the unknown filter parameters. These crude estimates can be used to estimate u_n as per (11) and hence $\Gamma(u_n)$ required for the second stage of *weighted* LS estimation. This procedure would eliminate convergence problems and much of the

computational complexity associated with the MLE. Yet it is apparent from (14), (17) and (19) that even if u_n were known computation of the weighting function might be difficult for many heavy-tailed non-Gaussian PDF's. Typically the weighting function involves computation of transcendental functions such as exponentials which may be computationally burdensome. The problem would be simplified considerably if Γ could be approximated by a simple function whose characteristics depended on the PDF parameters. A possible approximation of the weighting curves shown in Figures 1–3 is the *Butterworth "filter"*

$$\tilde{\Gamma}(u) = \frac{K_1}{1 + \left|\frac{u}{u_c}\right|^{\beta}} + K_2 \tag{20}$$

where u_c denotes the '3 dB cutoff', β is the order of approximation (not necessarily an integer) and $K_1 > 0$ and $K_2 > 0$ can be used to match $\tilde{\Gamma}$ with Γ for $u = 0$ and $u \to \infty$. All these parameters can be used to produce an accurate approximation of the usually complicated function Γ. The second stage of LS can therefore be simplified by minimizing (10) with Γ replaced by $\tilde{\Gamma}$. This yields

$$\sum_{i=1}^{p} a_i \left(\sum_{n=p+1}^{N} x_{n-i} x_{n-j} \tilde{\Gamma}(\hat{u}_n) \right) = - \sum_{n=p+1}^{N} x_n x_{n-j} \tilde{\Gamma}(\hat{u}_n), \qquad j = 1, 2, \cdots p$$

which in matrix form is

$$\underbrace{\begin{pmatrix} \sum_{n=p+1}^{N} x_{n-1}x_{n-1}\tilde{\Gamma}(\hat{u}_n) & \sum_{n=p+1}^{N} x_{n-1}x_{n-2}\tilde{\Gamma}(\hat{u}_n) & \cdots & \sum_{n=p+1}^{N} x_{n-1}x_{n-p}\tilde{\Gamma}(\hat{u}_n) \\ \sum_{n=p+1}^{N} x_{n-2}x_{n-1}\tilde{\Gamma}(\hat{u}_n) & \sum_{n=p+1}^{N} x_{n-2}x_{n-2}\tilde{\Gamma}(\hat{u}_n) & \cdots & \sum_{n=p+1}^{N} x_{n-2}x_{n-p}\tilde{\Gamma}(\hat{u}_n) \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{n=p+1}^{N} x_{n-p}x_{n-1}\tilde{\Gamma}(\hat{u}_n) & \sum_{n=p+1}^{N} x_{n-p}x_{n-2}\tilde{\Gamma}(\hat{u}_n) & \cdots & \sum_{n=p+1}^{N} x_{n-p}x_{n-p}\tilde{\Gamma}(\hat{u}_n) \end{pmatrix}}_{\mathbf{X}} \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{pmatrix}$$

$$= - \begin{pmatrix} \sum_{n=p+1}^{N} x_n x_{n-1}\tilde{\Gamma}(\hat{u}_n) \\ \sum_{n=p+1}^{N} x_n x_{n-2}\tilde{\Gamma}(\hat{u}_n) \\ \vdots \\ \sum_{n=p+1}^{N} x_n x_{n-p}\tilde{\Gamma}(\hat{u}_n) \end{pmatrix} \tag{21}$$

$\mathbf{X}$ is a symmetric $p \times p$ matrix which is positive semidefinite. To show this assume $\mathbf{b} = [b_1 \; b_2 \; \cdots \; b_p]$ is a vector of real numbers. Then,

$$\begin{aligned} \mathbf{b}^T\mathbf{X}\mathbf{b} &= \sum_{i=1}^{p}\sum_{j=1}^{p} b_i b_j \sum_{n=p+1}^{N} x_{n-i}x_{n-j}\tilde{\Gamma}(\hat{u}_n) \\ &= \sum_{n=p+1}^{N} \tilde{\Gamma}(\hat{u}_n) \sum_{i=1}^{p}\sum_{j=1}^{p} b_i b_j x_{n-i}x_{n-j} \\ &= \sum_{n=p+1}^{N} \tilde{\Gamma}(\hat{u}_n) \left[\sum_{i=1}^{p} b_i x_{n-i}\right]^2 \\ &\geq 0 \end{aligned}$$

since the weights $\tilde{\Gamma}(\hat{u}_n)$ are always positive (see (20)). (21) can be solved in $O(p^3)$ operations. A Cholesky decomposition can be used to reduce computation. This is in contrast to the unweighted least squares, for which it is possible to estimate the

parameters in $O(p^2)$ operations [Morf *et al* 1977]. In summary, the proposed technique is

STEP I : Use covariance method ((21) with $\tilde{\Gamma}(\hat{u}_n) = 1$) to obtain initial estimates of a.

STEP II : Generate the sequence $\hat{u}_n$ by passing $\{x_{p+1}, x_{p+2}, \cdots, x_N\}$ through the MA filter whose coefficients are as estimated in step I, as given by (11).

STEP III : Select the curve $\tilde{\Gamma}$ (20) by choosing appropriate values of u_c, β, K_1 and K_2 from the known values of the PDF parameters (Θ).

STEP IV : Solve for a from (21).

Step III will be different for different non-Gaussian PDF's. The following section addresses this part of the problem for the specific case of a *mixed-Gaussian* distribution.

IV. Weighted LS for Mixed-Gaussian PDF

Performance of the weighted LS estimator described in the previous section is expected to be dependent on how well the curve $\tilde{\Gamma}$ can approximate the true weighting curve Γ. Therefore the parameters of $\tilde{\Gamma}$, namely, u_c, β, K_1 and K_2 should be chosen properly for every set of values of the parameters of the PDF, *i.e.*, Θ. In the mixed-Gaussian case the parameters ϵ and ρ determine the shape of Γ (see Figures 1(a) and 1(b)). K_1 and K_2 can be found as a function of these two parameters by matching the values of Γ and $\tilde{\Gamma}$ for $u = 0$ and $u \to \infty$. It follows from (14) assuming $\sigma_B^2 = 1$ that

$$\Gamma(0) = \frac{(1-\epsilon) + \dfrac{\epsilon}{\rho\sqrt{\rho}}}{(1-\epsilon) + \dfrac{\epsilon}{\sqrt{\rho}}} \qquad \text{and} \qquad \Gamma(\infty) \to \frac{1}{\rho}$$

Also, $\tilde{\Gamma}(0) = K_1 + K_2$ and $\tilde{\Gamma}(\infty) \to K_2$. Therefore

$$K_1 = \left[\frac{(1-\epsilon) + \dfrac{\epsilon}{\rho\sqrt{\rho}}}{(1-\epsilon) + \dfrac{\epsilon}{\sqrt{\rho}}}\right] - \frac{1}{\rho} \qquad \text{and} \qquad K_2 = \frac{1}{\rho} \tag{22}$$

Fortunately, K_1 and K_2 are obtained as explicit closed-form functions of ϵ and ρ. σ_B^2 has been assumed to be unity without loss of generality, since it will only change u_c by a scale factor and furthermore the weighting curve need only be determined to within a scale factor for use in (21). u_c can be chosen to match $\tilde{\Gamma}$ and Γ at $u = u_c$. It is found by solving

$$\Gamma(u_c) = \frac{K_1}{2} + K_2 \tag{23}$$

where Γ and is defined by (14). Γ being a complicated function, (23) can only be solved by a search algorithm. Figure 4 plots u_c, as obtained by solving (23), vs. ϵ for different values of ρ. The curves can be explained by interpreting the mixed-Gaussian PDF as one arising from a nominally Gaussian noise contaminated by a high variance Gaussian process. A small value of ϵ implies little contamination by the high variance population and therefore the *limiter* can allow for reasonably large values of u_n and hence a large u_c results. On the other hand as ϵ increases, increased interference from the high variance population is compensated for by making the threshold u_c smaller so as not to allow large values of u_n suppress the information contained in the other terms. Figure 5 plots u_c vs. ρ for different values of ϵ. It shows that the threshold is minimum for $\rho \approx 10$. If ρ is large, the contaminating population would introduce very large spikes and therefore a high threshold would suffice to suppress them. For a smaller ratio of σ_I^2 to σ_B^2 a smaller threshold is necessary. When σ_I^2 and σ_B^2 are of the same order ($\rho < 10$), it becomes difficult to distinguish between contributions from the two populations and therefore most of the terms should be equally weighted, which is accomplished by causing the threshold to be large, as can be verified from Figure 5. An interesting special case is $\rho = 1$ when the mixed-Gaussian PDF degenerates to a Gaussian PDF ($\Gamma(u_n) = 1/\sigma_B^2$ for all σ_B^2). The threshold goes to ∞ and all the terms are equally weighted. Once the threshold is calculated, the most appropriate β for a given ϵ and ρ can be found by a least squares curve fitting method. Specifically, a suitable range of u (where $\Gamma(u)$ is significantly positive) is divided in 1000 equally

spaced points. The sum of $(\Gamma(u) - \tilde{\Gamma}(u))^2$ evaluated at these points is then minimized with respect to β. Figure 6 plots β vs. ϵ for different values of ρ. It shows that a sharp cutoff (*i.e.*, a high value of β) is necessary only when the contaminating process has high power and it appears very rarely (large ρ and small ϵ).

A typical approximation of $\Gamma(u)$ by $\tilde{\Gamma}(u)$ has been plotted in Figure 7. Both the functions are plotted vs. u in the same scale. $\sigma_B^2 = 1$, $\rho = 100$ and $\epsilon = 0.1$ were assumed. The corresponding threshold u_c was 3.0224 and the most suitable β was 9.422. $K_1 = 0.979$ and $K_2 = 0.01$ were obtained from (22). The approximation appers to be reasonably accurate. In general, accuracy of the approximation will depend on the values of ρ and ϵ.

V. The case of unknown PDF parameters

It was assumed for the weighted LS estimator described in section III that Θ, the vector of noise PDF parameters was known. This was necessary to in order to determine the weighting function to be used in (21). If it is partially or completely unknown, it has to be estimated. This will undoubtedly degrade the performance of the estimator. It will be shown in the next section that when the PDF parameters are known, the estimator performance of the weighted LS estimator proposed nearly attains the CR bound. Alternately, the performance is as good as the MLE. Hence for the purpose of discussion the weighted LS estimator can be considered to be asymptotically efficient when the PDF parameters are known so that Γ is known. It is thus of interest to determine the sensitivity of this performance to changes in Γ due to estimation errors in the unknown PDF parameters. One way of quantifying this is to determine the efficiency of the corresponding AR filter parameter estimator as the PDF parameters vary from the true values assumed for Γ. The problem is now examined from the viewpoint of robust M-estimators [Martin 1979], [Martin and Yohai 1984].

The original set of non-linear equations (6) to be solved for the MLE (which are

approximated by a set of linear equations in the weighted LS method) can be written as

$$\sum_{n=p+1}^{N} x_{n-j}\phi\left(x_n + \sum_{i=1}^{p} a_i x_{n-i}\right) = 0, \qquad j = 1, 2, \cdots p \tag{24}$$

where $\phi(u_n) = u_n\Gamma(u_n) = -f'(u_n;\Theta)/f(u_n;\Theta$ is an odd function. If $\Gamma(u_n)$ is as defined in (5) with the true PDF parameters then the solution of (24), if it exists and is the unique maximum of the likelihood function, is the MLE of a (assuming known PDF parameters). If $\Gamma(u_n)$ is replaced by a different limiter curve $\tilde{\Gamma}(u_n)$, the resulting estimator is termed an M-estimator [Huber 1981]. When the true PDF is not perfectly known, ϕ (or $\tilde{\Gamma}$) has to be selected on the basis of other considerations, *e.g.*, making the estimator performance less sensitive to the PDF. The performance of an estimator so designed is not as good as the MLE (which is based on the perfect knowledge of the PDF), but assuming $\tilde{\Gamma}$ is well chosen its performance does not deteriorate much if the actual PDF is somewhat different from *the PDF which produces best performance for a particular selection of* ϕ. Such an estimator exhibits *efficiency robustness* if the performance is evaluated in terms of asymptotic efficiency. The asymptotic efficiency of an estimator of a can be quantified by [Anderson 1971]

$$EFF(\hat{\mathbf{a}}, f) = \left[\frac{\det I_{aa}^{-1}(\mathbf{a})}{\det V(\hat{\mathbf{a}})}\right]^{\frac{1}{p}} \tag{25}$$

where $I_{aa}(\mathbf{a})$ is the information matrix for $\mathbf{a}$ and $V(\hat{\mathbf{a}})$ is the *asymptotic* covariance matrix of $\hat{\mathbf{a}}$. It can be shown that $0 \leq EFF(\hat{\mathbf{a}}, f) \leq 1$ and the upper bound is reached if and only if $V(\hat{\mathbf{a}}) = I_{aa}^{-1}(\mathbf{a})$. In the case of M-estimates, assuming the PDF f is symmetric, (25) reduces to [Martin 1979]

$$EFF(\hat{\mathbf{a}}, f) = \frac{\left[E[\phi']\right]^2}{I_f E[\phi^2]} \tag{26}$$

where

$$I_f = E\left[\left(\frac{f'}{f}\right)^2\right] \tag{27}$$

The asymptotic efficiency given by (26) depends on how well the function ϕ matches the optimal one for a given PDF f. It attains the upper bound of unity if only if $\phi = -f'/f$ or alternately (24) represents the MLE equations.

For a *Gaussian* M-estimator (which *assumes* the underlying PDF to be Gaussian with zero mean and variance σ^2 for the purpose of choosing ϕ) $\phi = -f'/f = u_n/\sigma^2$. Therefore the estimator reduces to a LS estimator. In this case, assuming the true PDF to be f, (26) implies

$$EFF(\hat{a}, f) = \frac{1}{\sigma^2 I_f} \tag{28}$$

It is known [Sengupta and Kay 1986] that for all symmetric PDF's $\sigma^2 I_f \geq 1$ with the equality holding *only* for the Gaussian PDF. Therefore for all symmetric non-Gaussian PDF's efficiency of the LS estimator is less than unity. In fact the LS estimator is known to be severely lacking in efficiency robustness, as verified by Figures 8, 9, 10 and 11 which plot the asymptotic efficiency of the LS estimator given by (28) for the three non-Gaussian PDF's described in Section II. A small deviation from Gaussianity is observed to produce a large drop in asymptotic efficiency. For example, in the case of a mixed-Gaussian PDF, it can be observed from Figure 8 that $\rho = 100$ and $\epsilon = 0.1$ results in a drop by a factor of 10 in the asymptotic efficiency from the value at $\epsilon = 0$ (which corresponds to a Gaussian PDF). Figure 9, which plots the asymptotic efficiency of the LS estimator for a mixed-Gaussian process vs. ρ for different values of ϵ, shows that the estimator loses efficiency for moderately large values of ρ, even when ϵ is reasonably small. In the cases of Middleton's class A and Johnson's families Gaussianity corresponds to large values of B and t, respectively. In both cases the asymptotic efficiency of the LS estimator drops substantially when these parameters are smaller (see Figures 10 and 11).

It is expected that a wiser choice of ϕ (or Γ) in (24) would result in a better M-estimator. If the PDF parameters (Θ) were known, the choice $\Gamma = -f'/uf$ or a suitable approximation $\tilde{\Gamma}$ thereof would have been optimal. Since these parameters

are unknown, a selection of $\tilde{\Gamma}$ which is quite appropriate for one value of Θ may not be suitable for other values of it. If such a mismatch does not reduce the asymptotic efficiency *substantially*, the corresponding M-estimator would be considered insensitive to small variations in the PDF parameters. This will be examined by plotting the asymptotic efficiency of a nominal M-estimator as the true PDF parameter values vary from the values for which the chosen estimator is optimal. The weighted LS estimator proposed in section III may be viewed as an M-estimator *where* u_n *(the argument of* $\tilde{\Gamma}$*) is replaced by a preliminary estimate* $\hat{u}_n$. Therefore the asymptotic performance (in terms of efficiency) of the M-estimators, which will now be described, should be a good indication of the sensitivity of the weighted LS estimator to changes in $\tilde{\Gamma}$ due to estimation errors in unknown PDF parameters.

Figure 12 plots the asymptotic efficiency of a typical M-estimator for different mixed-Gaussian PDF's. A fixed limiter curve with $u_c = 3$, $\beta = 10$, $K_1 = 0.98$ and $K_2 = 0.01$ is used. In this case

$$\phi(u_n) = u_n\tilde{\Gamma}(u_n) = \frac{0.98u_n}{1 + \left|\frac{u_n}{3}\right|^{10}} + 0.01u_n \tag{29}$$

σ_B^2 is assumed to be unity and the asymptotic efficiency (calculated from (26) and (29) by numerical integration) is plotted vs. ϵ for different values of ρ. The curve corresponding to $\rho = 100$ exhibits a maximum (efficiency ≈ 1) at $\epsilon = 0.1$ demonstrating that a mixed-Gaussian PDF with $\epsilon = 0.1$ and $\rho = 100$ is most suitable for this M-estimator. In fact, these values of ϵ and ρ were *actually* used to determine $\tilde{\Gamma}$ as described in Section IV and resulted in (29). For values of ϵ from 0 to 0.5 its asymptotic performance is reasonably good. The asymptotic efficiency is 0.98 at $\epsilon = 0$ (*i.e.*, the PDF is Gaussian) and 0.90 at $\epsilon = 0.5$. Therefore for a truly Gaussian PDF this estimator will be 98% as efficient as a LS estimator (which has efficiency of one for a Gaussian PDF). This can be thought of as a 2% *premium* [Martin 1979] for a 90% protection or *coverage* against up to 50% outliers. Figure 13 plots the asymptotic efficiency of the same estimator vs.

ρ for different values of ϵ. It shows that although this particular choice of ϕ is most suitable for $\epsilon = 0.1$ and $\rho = 100$, the asymptotic efficiency is more than 90% up to $\rho = 10000$ for values of ϵ less than 0.1. Improvement over the curves of Figure 9 is quite apparent.

ϵ and ρ were assumed to be known in the derivation of the weighted LS estimator. Figures 12 and 13 show that the asymptotic efficiency of the corresponding M-estimator is rather insensitive to these parameters. Assuming that the result extends to the weighted LS estimator, it implies that when these parameters are *not* known, they can be approximated by crude estimates for the purpose of selecting $\tilde{\Gamma}$ in step III of the suggested estimation procedure. To be more precise, the parameters of $\tilde{\Gamma}$ (namely, u_c, β, K_1 and K_2) can be stored in a table as functions of the unknown PDF parameters and the proper values can be chosen by interpolation from these tables. Hardware memories can be used for this purpose for on-line estimation. The resulting estimator will be adaptive in nature because it would select a limiter curve depending on a crude estimate of the unknown PDF parameters. This result adds flexibility to the method and also creates a possibility of estimating the unknown PDF parameters more precisely once the AR filter parameters are estimated accurately.

The mixed-Gaussian PDF is not the only PDF which provides such an opportunity. Figures 14 and 15, which plot the asymptotic efficiency of the M-estimator (calculated with a typical selection of $\tilde{\Gamma}$ in each case) for Middleton's class A and Johnson families, suggest the existence of similar results in the cases of other non-Gaussian PDF's. The parameters of $\tilde{\Gamma}$ chosen for Figure 13 were $u_c = 1.8$, $\beta = 7$, $K_1 = 0.955$ and $K_2 = 0.045$, which are quite suitable for Middleton's class A PDF with $A = 0.05$ and $B = 1.0$. For smaller values of B the M-estimator shows marked improvement over the LS estimator (compare Figure 10). The parameters of $\tilde{\Gamma}$ for the Johnson's family were chosen to be $u_c = 1.4$, $\beta = 2$, $K_1 = 0.98$ and $K_2 = 0.02$. These are suitable for Johnson's PDF with $t = 1.5$. A comparison of Figures 11 and 15 reveals that the M-estimator

does not lose efficiency as fast as the LS estimator as t becomes smaller (*i.e.* the PDF becomes more non-Gaussian). These are instances of the M-estimator being insensitive to small variations in some of the PDF parameters. It is not clear how well these results apply to the weighted LS estimator, and hence its improvement over the LS estimator may be somewhat less than what the curves show. The central argument is that the proposed estimator improves the efficiency robustness of the LS estimator by weighting the squared terms, and it also reduces computation over the M-estimator by approximating the argument of $\tilde{\Gamma}$. Its ability to handle the case of *unknown PDF parameters* makes it more attractive in practice than an M-estimator which requires the solution of non-linear equations. The following section presents the results of computer simulations which justify the approximations made in deriving the weighted LS estimator.

VI. Simulation of the performance of the weighted LS estimator

Two typical AR(4) processes [Kay 1986] was chosen for computer simulations. The parameters are given in Table A. Process I is broadband while process II is narrowband. The underlying PDF is assumed to be mixed-Gaussian with $\sigma_B^2 = 1$ and $\rho = 100$. The mixture parameter is $\epsilon = 0.1$. The AR process was generated by passing a white mixed-Gaussian process through a filter, allowing sufficient time for the transients to decay. The white process was generated by randomly selecting from two mutually independent white Gaussian processes with PDF's E_B and E_I (having variances σ_B^2 and $\sigma_I^2 = \rho\sigma_B^2$ respectively) on the basis of a series of Bernoulli trials with probability of success ϵ. Thus a random variable could be expected to come from the *background* population for (1-ϵ) fraction of times and from the *contaminating* population for ϵ fraction of times. In accordance with the discussion in the previous section one of the PDF parameters, namely ϵ, was assumed to be unknown. ϵ is linearly related to the overall variance σ^2

of the PDF,

$$\sigma^2 = \sigma_B^2[(1-\epsilon) + \epsilon\rho)] \tag{30a}$$

i.e.,

$$\epsilon = \frac{1}{\rho - 1}\left[\frac{\sigma^2}{\sigma_B^2} - 1\right] \tag{30b}$$

It was suggested in the previous section that a crude estimator of ϵ can be used to select the proper weighting curve. In this case the driving noise power, *i.e.*, σ^2 was estimated along with a in the first step of unweighted LS estimation using covariance method (see (8)) and ϵ was calculated from this estimate using (30b).

Table B shows the sample means and sample variances of the AR filter parameter estimators obtained by the exact evaluation of the approximate MLE. The MLE is found by the four-dimensional optimization (for the four AR filter parameters) of the conditional likelihood function as reported in [Sengupta and Kay 1986]. A Newton-Raphson iterative procedure was used for this purpose, with initial conditions obtained from a preliminary stage of least squares estimation, namely, the Forward-backward method [Kay 1986]. The value of ϵ used in (6) was as obtained from σ^2 in the first stage of LS estimation. 1000 data points were used and the result is based on 500 experiments. The results, as summarized in Table D, can be compared to the performance of the Forward-Backward estimator (see Table C) and the CR bound. The MLE achieves the CR bound while the variances of the Forward-Backward estimators are larger by a factor of 10. This agrees with the theoretical prediction of previous section, as the asymptotic efficiency of an LS estimator, given in this case by (28), is $0.106 \approx 1/10$. Table D reports the performance of the weighted LS estimator. The loss of performance is only marginal, as is verified by comparing it to the performance of the exact MLE.

Evaluation of the exact MLE is not only computationally intensive, but it also suffers from convergence problems. For 1000 data points 1% of the experiments failed

to converge. For short data records (*e.g.*, $N < 250$) it almost never converges. However the weighted LS does not suffer from this problem. The performance of the unweighted and weighted LS estimators for $N = 100$ is summarized in Tables E and F, respectively, along with the CR bound. The unweighted LS (Forward-Backward) estimator continues to be off from the CR bound by a factor of 10 for Process I. The offset increases to a factor of 15 for Process II. The weighted LS suffers from a slight degradation of performance: it is off from the CR bound by a factor of 2 for Process I and by a factor of 2.5 for Process II. This is probably due to the increased inaccuracy of the simplifying approximations for shorter data records. It still exhibits improvement over the Forward-Backward estimator. It is also seen to have less bias as compared to the Forward-Backward estimator in all cases.

VII. Conclusions

The weighted LS estimator proposed in this paper yields accurate estimates of the parameters of an AR process excited by non-Gaussian white noise. The method utilizes the partial information available about the noise PDF (principally the form of the PDF to within a set of unknown parameters) and should thereby outperform the so-called *robust* estimators. It also reduces computation by avoiding solution of non-linear equations required by the MLE or a robust estimator. Computer simulations have justified the assumptions made in determining the estimator. The new technique does not suffer from convergence problems, and exhibits only a slight departure in performance from the CR bound for short data records. The weighted LS estimator for the AR filter parameters can be used in conjunction with other estimation techniques directed towards assessment of unknown PDF parameters. In some situations it may tolerate a reasonable inaccuracy in the estimation of these parameters and yet produce an accurate estimate of the AR filter parameters.

References

[1] S.M. Kay, *Modern Spectral Estimation: Theory and Application*, Chapters 5-7, Prentice-Hall, 1986, to be published.

[2] D. Sengupta and S.M. Kay, "Efficient Estimation of Parameters for Non-Gaussian Autoregressive Processes", submitted to IEEE Trans. on Acoustics, Speech and Signal Processing, 1986.

[3] R.D. Martin, "The Cramer-Rao Bound and Robust M-Estimates for Time Series Autoregressions", Biometrica, pp. 437–442, Vol. 69, No. 2, 1982.

[4] D. Sengupta, "Estimation and Detection for Non-Gaussian Processes Using Autoregressive and Other Models", M.S. Thesis, Univ. of Rhode Island, 1986.

[5] S.L. Bernstein *et al.*, "Long-range Communication at extremely low frequencies", Proc. of the IEEE, pp. 292–312, Vol. 62, March 1974.

[6] G.E.P. Box and G.J. Jenkins, *Time Series Analysis: Forecasting and Control*, Chapter 7, San Francisco: Holden-Day, 1970.

[7] S.V. Czarnecki and J.B. Thomas, "Nearly Optimal Detection of Signals in Non-Gaussian noise", ONR report #14, Feb. 1984.

[8] D. Middleton, "Cannonically Optimum Threshold Detection", IEEE Trans. on Info. Theory, pp. 230–243, Vol. IT-12, No. 2, April 1966.

[9] N.L. Johnson and S. Kotz, *Distributions in Statistics: Continuous Univariate Distributions*, New York: John Wiley, 1971.

[10] R.D. Martin, "Robust Estimation for Time Series Autoregressions" in *Robustness in Statistics*, Edited by R.L. Launer and G.N. Wilkinson, New York: Academic, 1979.

[11] R.D. Martin and V.J. Yohai, "Robustness in Time Series and Estimating ARMA models", Univ. of Washington Technical Report #50, June 1984.

[12] P.J. Huber, *Robust Statistics*, Chapter 4, New York: John Wiley, 1981.

[13] T.W. Anderson, *The Statistical Analysis of Time Series*, Chapter 5, New York: John Wiley, 1971.

[14] H.W. Sorenson, *Parameter Estimation: Principles and Problems*, Chapter 2, New York: Marcel Dekker, 1980.

[15] L. Pakula, private communication, 1986.

Table A: Parameters of the AR processes used for simulation

Process	a_1	a_2	a_3	a_4	poles
I	-1.352	1.338	-0.662	0.240	$0.7\exp[j2\pi(0.12)]$ $0.7\exp[j2\pi(0.21)]$
II	-2.760	3.809	-2.654	0.924	$0.98\exp[j2\pi(0.11)]$ $0.98\exp[j2\pi(0.14)]$

Table B: Performance of the MLE, $N = 1000$

		True value	Sample mean	Bias^2	Sample variance	Cramer-Rao bound
Process I	a_1	-1.352	-1.3527	4.900×10^{-7}	1.0221×10^{-4}	1.0491×10^{-4}
	a_2	1.338	1.3391	1.210×10^{-6}	2.4601×10^{-4}	2.5961×10^{-4}
	a_3	-0.662	-0.6630	1.000×10^{-6}	2.4251×10^{-4}	2.5961×10^{-4}
	a_4	0.240	0.2404	1.600×10^{-7}	1.1035×10^{-4}	1.0491×10^{-4}
Process II	a_1	-2.760	-2.7597	9.000×10^{-8}	1.7445×10^{-5}	1.6278×10^{-5}
	a_2	3.809	3.8081	8.100×10^{-7}	9.0097×10^{-5}	8.0163×10^{-5}
	a_3	-2.654	-2.6530	1.000×10^{-6}	9.1303×10^{-5}	8.0163×10^{-5}
	a_4	0.924	0.9236	1.600×10^{-7}	1.8496×10^{-5}	1.6278×10^{-5}

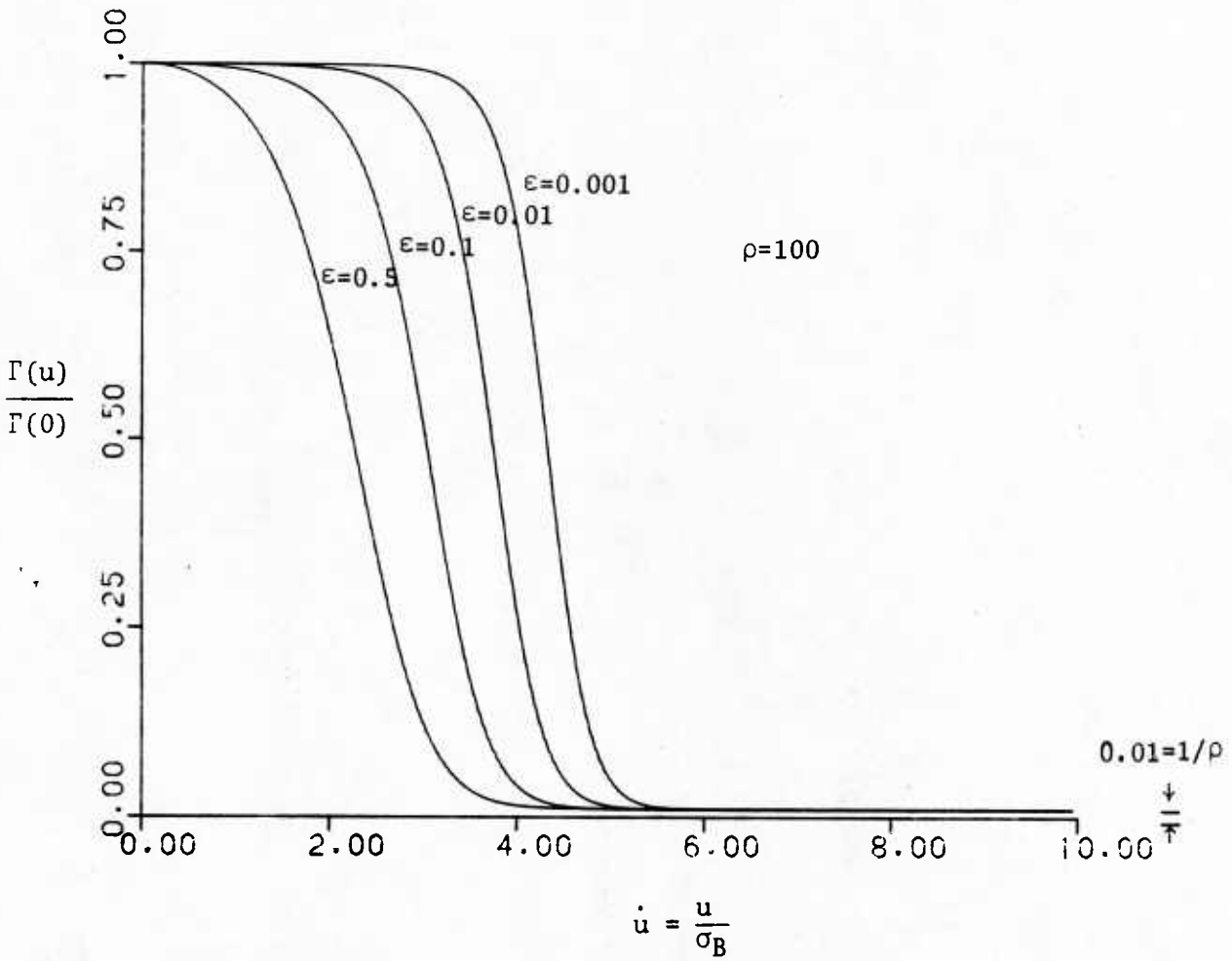

Figure 1(a) The weighting curve for Mixed-Gaussian PDF for different ε's

Table C: Performance of the Forward-Backward Estimator, $N = 1000$

		True value	Sample mean	Bias^2	Sample variance	Cramer-Rao bound
Process I	a_1	-1.352	-1.3482	1.444×10^{-5}	1.0197×10^{-3}	1.0491×10^{-4}
	a_2	1.338	1.3326	2.916×10^{-5}	2.3822×10^{-3}	2.5961×10^{-4}
	a_3	-0.662	-0.6591	8.410×10^{-6}	2.3531×10^{-3}	2.5961×10^{-4}
	a_4	0.240	0.2382	3.240×10^{-4}	9.6246×10^{-4}	1.0491×10^{-4}
Process II	a_1	-2.760	-2.7567	1.089×10^{-5}	1.6569×10^{-4}	1.6278×10^{-5}
	a_2	3.809	3.8001	7.921×10^{-5}	8.3418×10^{-4}	8.0163×10^{-5}
	a_3	-2.654	-2.6447	8.649×10^{-5}	8.4388×10^{-4}	8.0163×10^{-5}
	a_4	0.924	0.9197	1.849×10^{-5}	1.7083×10^{-4}	1.6278×10^{-5}

Table D: Performance of the Weighted LS Estimator, $N = 1000$

		True value	Sample mean	Bias^2	Sample variance	Cramer-Rao bound
Process I	a_1	-1.352	-1.3525	2.500×10^{-7}	1.1169×10^{-4}	1.0491×10^{-4}
	a_2	1.338	1.3388	6.400×10^{-7}	2.6466×10^{-4}	2.5961×10^{-4}
	a_3	-0.662	-0.6628	6.400×10^{-7}	2.6389×10^{-4}	2.5961×10^{-4}
	a_4	0.240	0.2402	4.000×10^{-8}	1.1049×10^{-4}	1.0491×10^{-4}
Process II	a_1	-2.760	-2.7595	2.500×10^{-7}	1.9003×10^{-5}	1.6278×10^{-5}
	a_2	3.809	3.8075	2.250×10^{-6}	9.8729×10^{-5}	8.0163×10^{-5}
	a_3	-2.654	-2.6524	2.560×10^{-6}	1.0043×10^{-4}	8.0163×10^{-5}
	a_4	0.924	0.9233	4.900×10^{-7}	2.0422×10^{-5}	1.6278×10^{-5}

Table E: Performance of the Forward-Backward Estimator, $N = 100$

		True value	Sample mean	Bias^2	Sample variance	Cramer-Rao bound
Process I	a_1	-1.352	-1.3328	3.686×10^{-4}	9.7580×10^{-3}	1.0491×10^{-3}
	a_2	1.338	1.3095	8.123×10^{-4}	2.1536×10^{-2}	2.5961×10^{-3}
	a_3	-0.662	-0.6383	5.617×10^{-4}	2.0595×10^{-2}	2.5961×10^{-3}
	a_4	0.240	0.2324	5.776×10^{-5}	8.5016×10^{-3}	1.0491×10^{-3}
Process II	a_1	-2.760	-2.7372	5.198×10^{-4}	2.4807×10^{-3}	1.6278×10^{-4}
	a_2	3.809	3.7494	3.552×10^{-3}	1.2929×10^{-2}	8.0163×10^{-4}
	a_3	-2.654	-2.5934	3.672×10^{-3}	1.2943×10^{-2}	8.0163×10^{-4}
	a_4	0.924	0.8974	7.076×10^{-4}	2.5026×10^{-3}	1.6278×10^{-4}

Table F: Performance of the Weighted LS Estimator, $N = 100$

		True value	Sample mean	Bias^2	Sample variance	Cramer-Rao bound
Process I	a_1	-1.352	-1.3476	1.936×10^{-5}	1.9844×10^{-3}	1.0491×10^{-3}
	a_2	1.338	1.3293	7.569×10^{-5}	5.2554×10^{-3}	2.5961×10^{-3}
	a_3	-0.662	-0.6536	7.056×10^{-5}	5.2139×10^{-3}	2.5961×10^{-3}
	a_4	0.240	0.2366	1.156×10^{-5}	2.0403×10^{-3}	1.0491×10^{-3}
Process II	a_1	-2.760	-2.7543	3.249×10^{-5}	3.6109×10^{-4}	1.6278×10^{-4}
	a_2	3.809	3.7936	2.372×10^{-4}	1.9891×10^{-3}	8.0163×10^{-4}
	a_3	-2.654	-2.6380	2.560×10^{-4}	2.0666×10^{-3}	8.0163×10^{-4}
	a_4	0.924	0.9168	5.184×10^{-5}	4.3322×10^{-4}	1.6278×10^{-4}

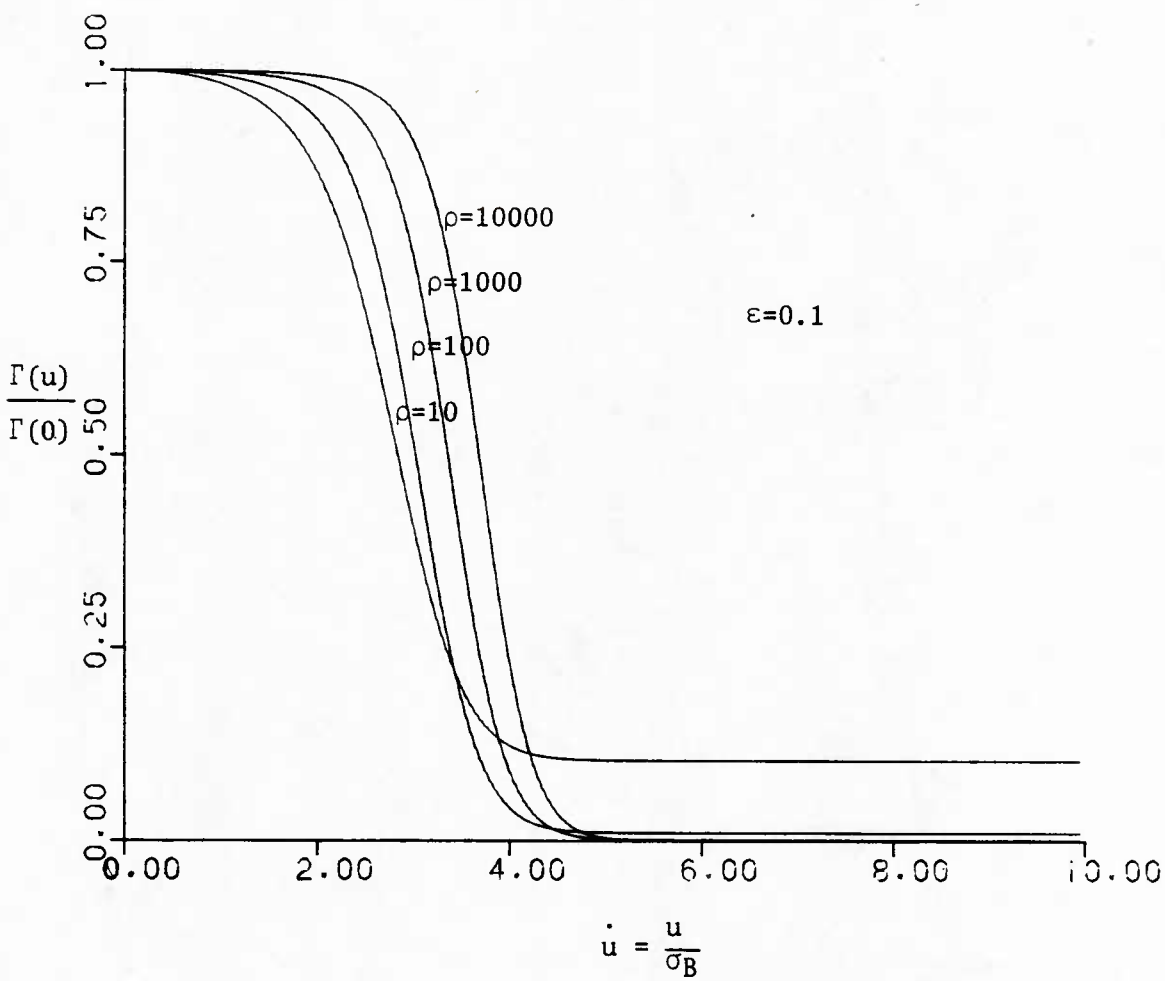

Figure 1(b) The weighting curve for Mixed-Gaussian PDF for different 's

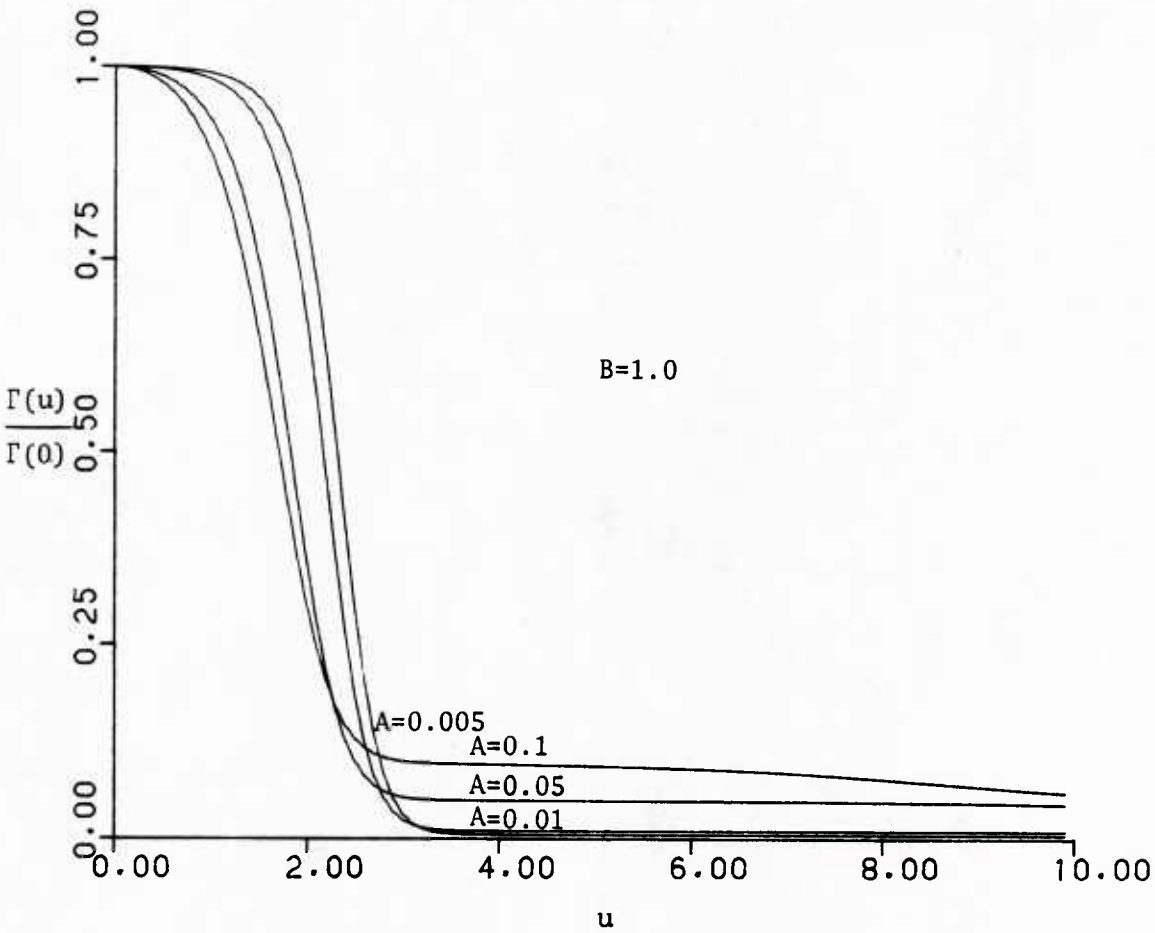

Figure 2(a) The weighting curve for Middleton's class A PDF for different A's

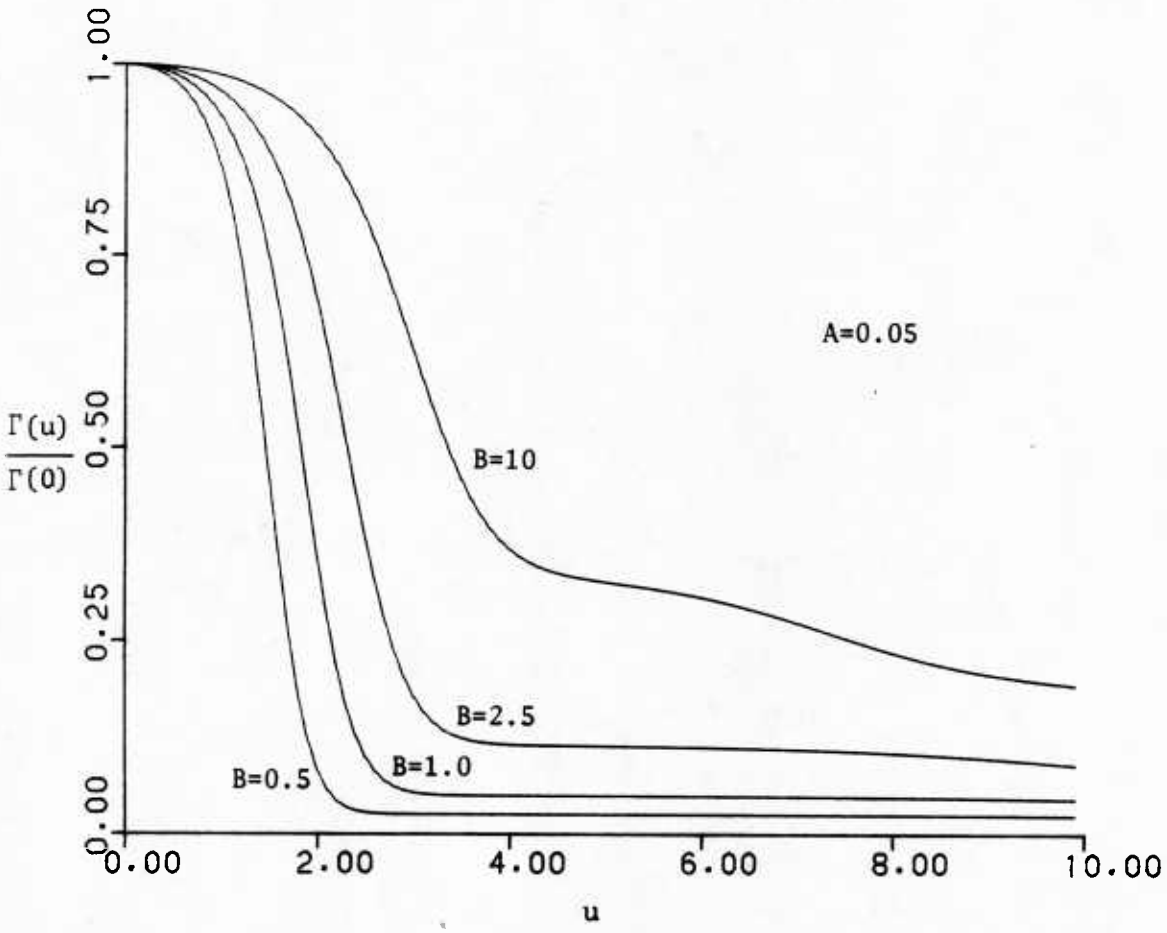

Figure 2(b) The weighting curve for Middleton's class A PDF for different B's

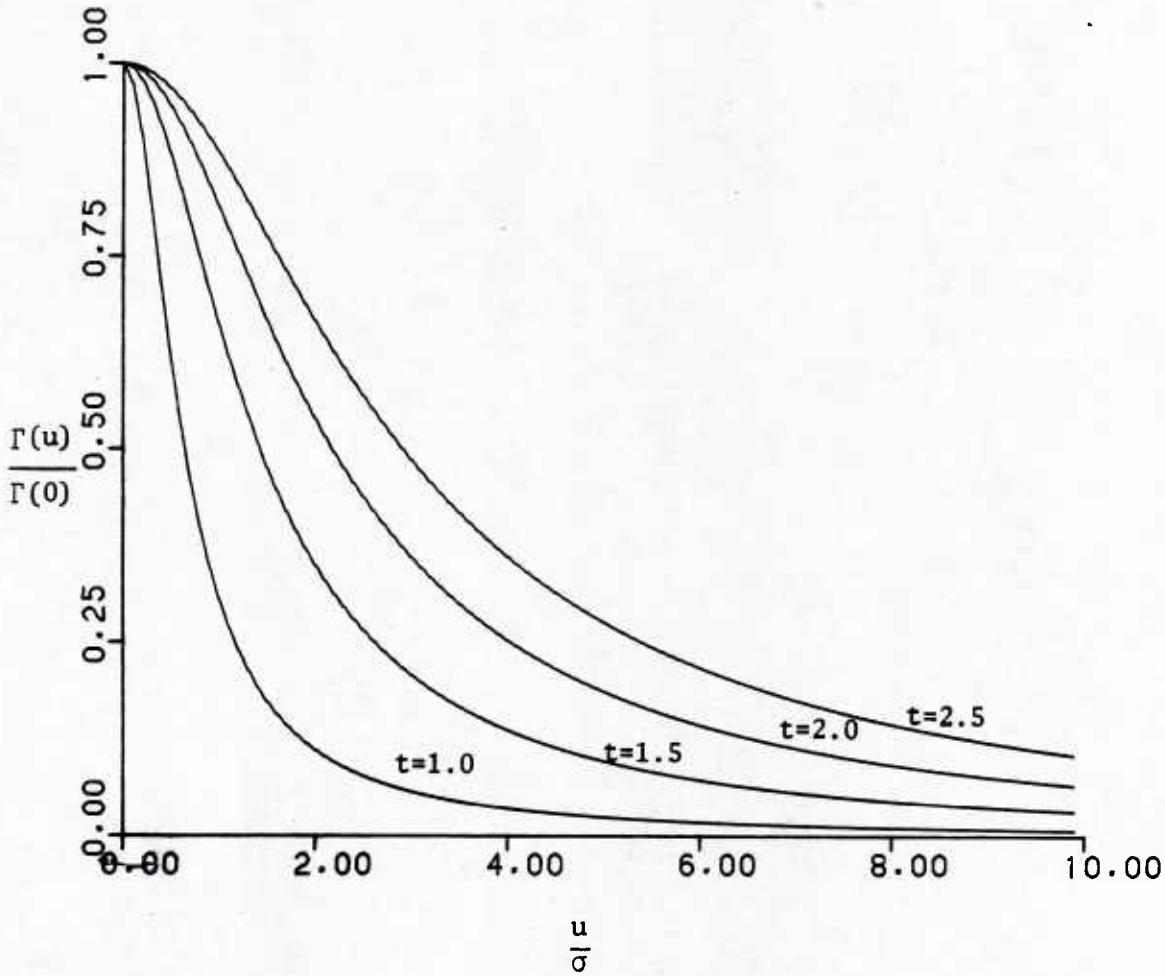

Figure 3 The weighting curve for Johnson's PDF for different t's

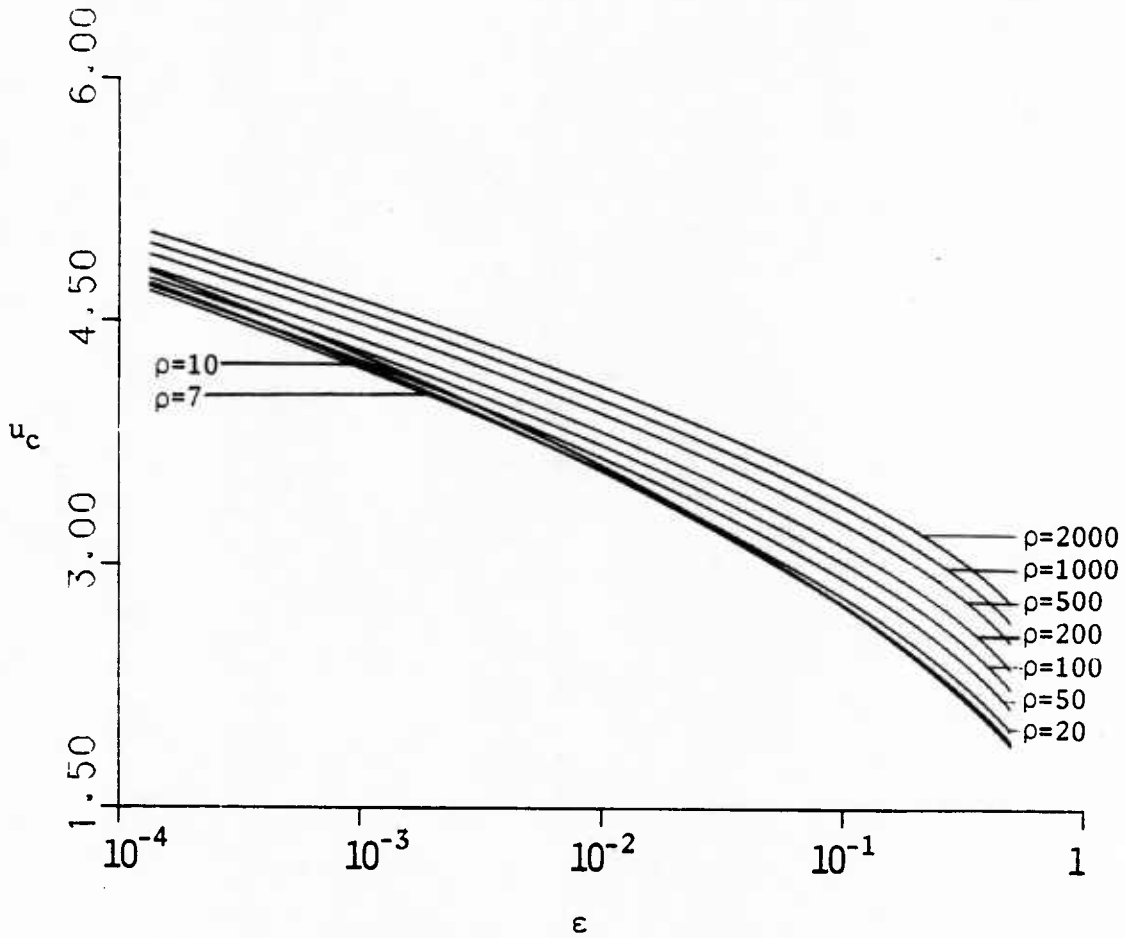

Figure 4 Dependence of threshold on ε in the Mixed-Gaussian case

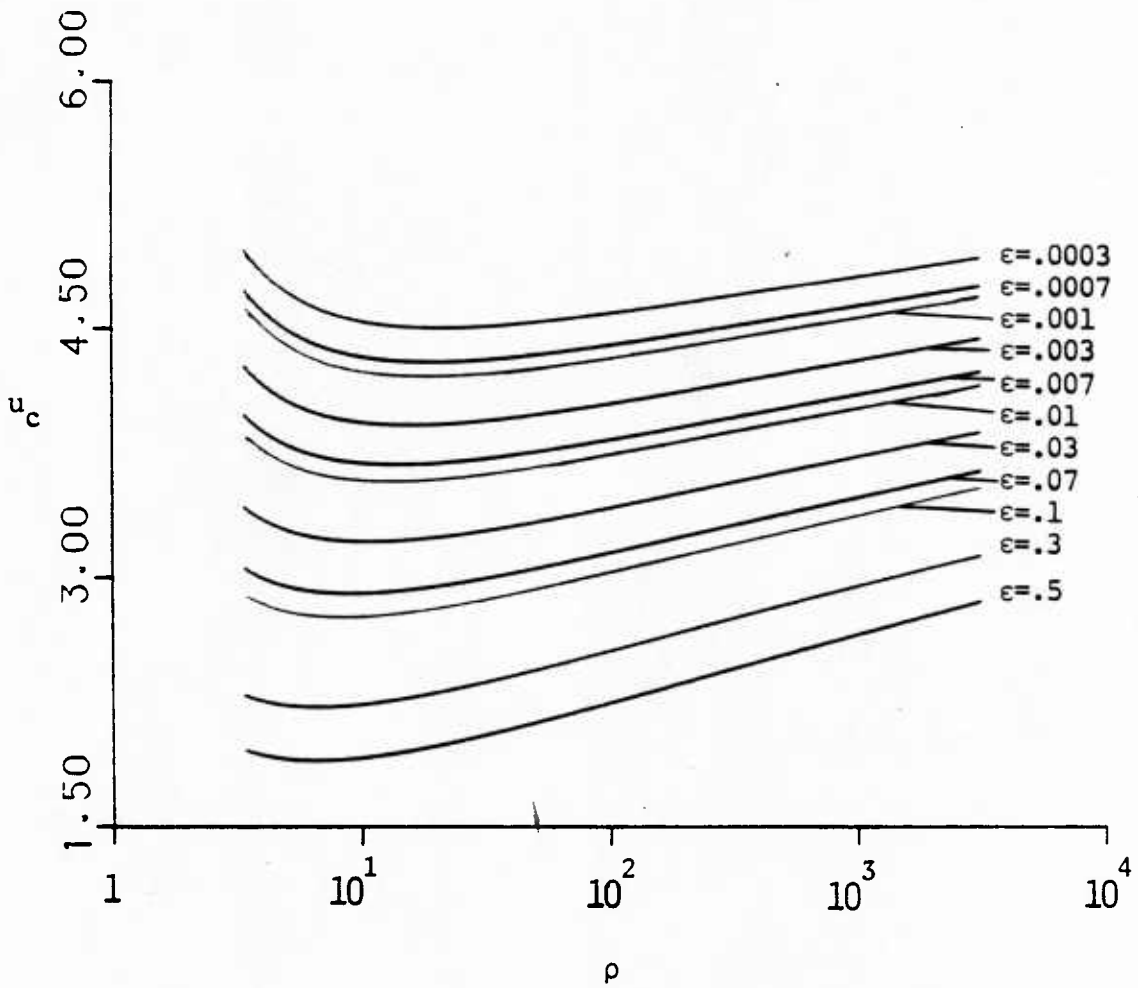

Figure 5 Dependence of threshold on ρ
in the Mixed-Gaussian case

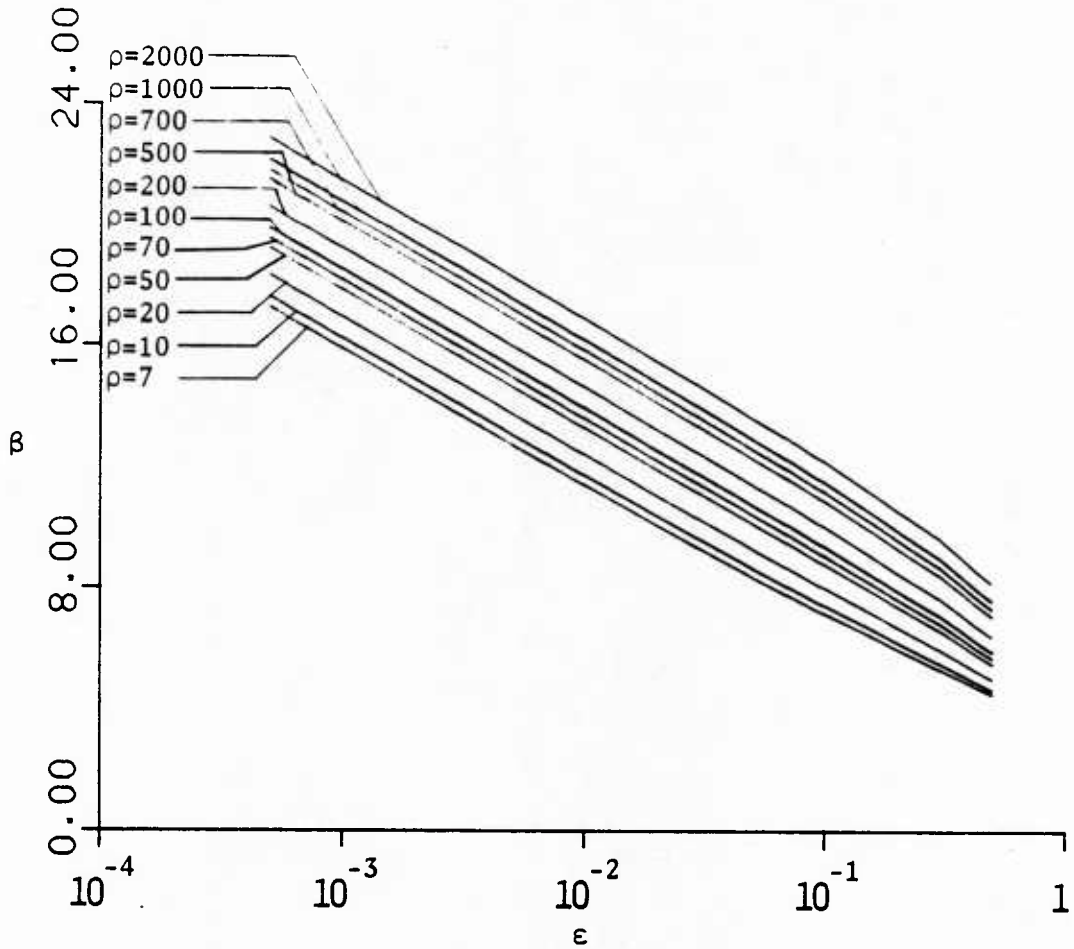

Figure 6 Dependence of β on ε and ρ in the Mixed-Gaussian case

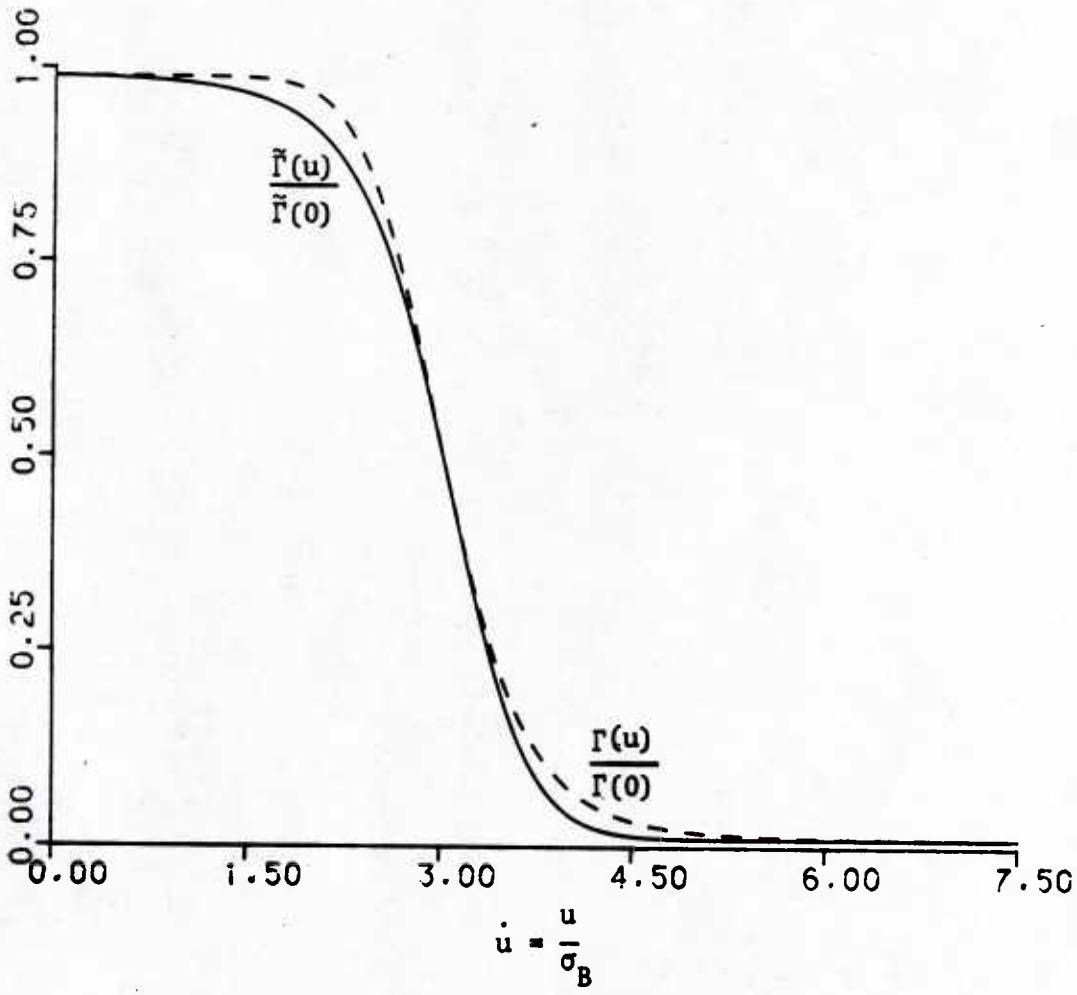

Figure 7 A typical approximation for the weighting curve in the Mixed-Gaussian case

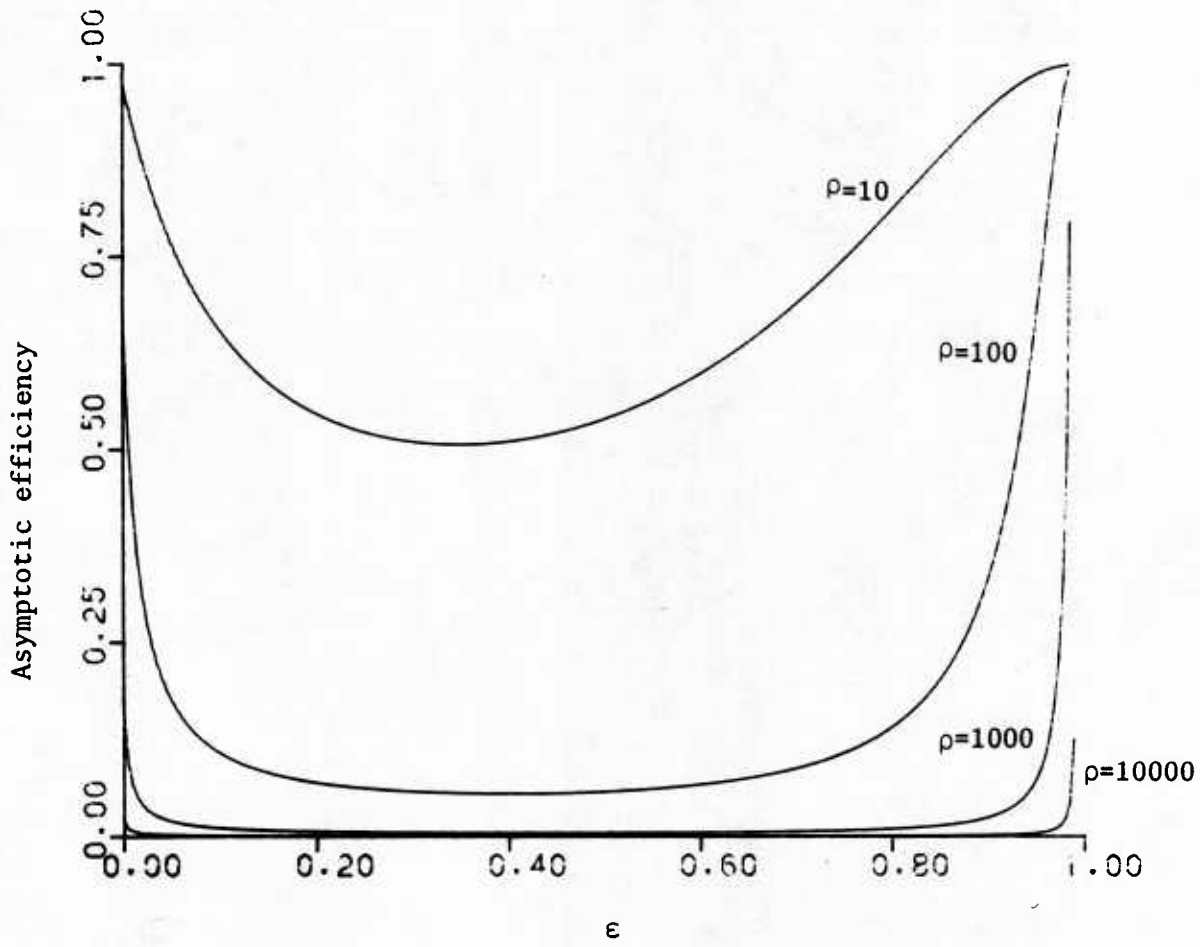

Figure 8 Asymptotic efficiency of LS estimator vs. ε for the Mixed-Gaussian process

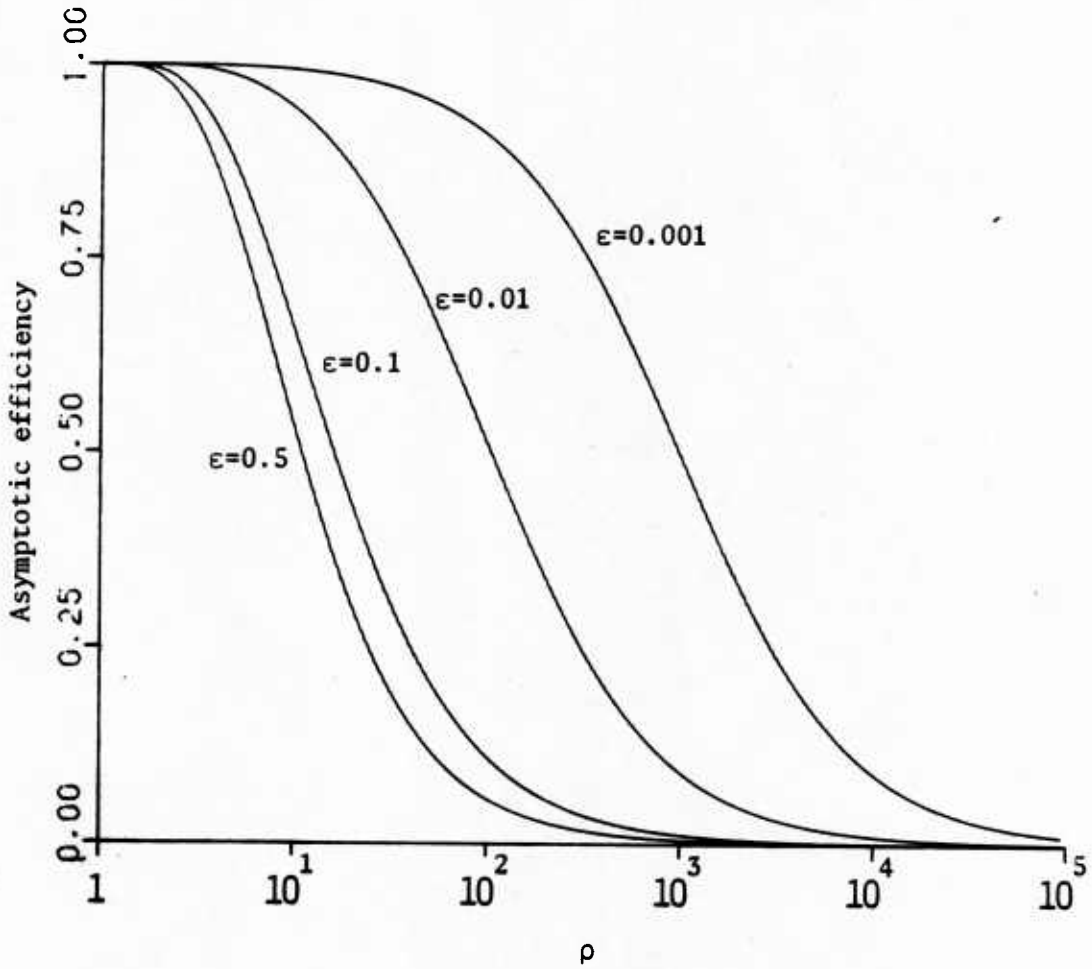

Figure 9 Asymptotic efficiency of LS estimator vs. ρ for the Mixed-Gaussian process

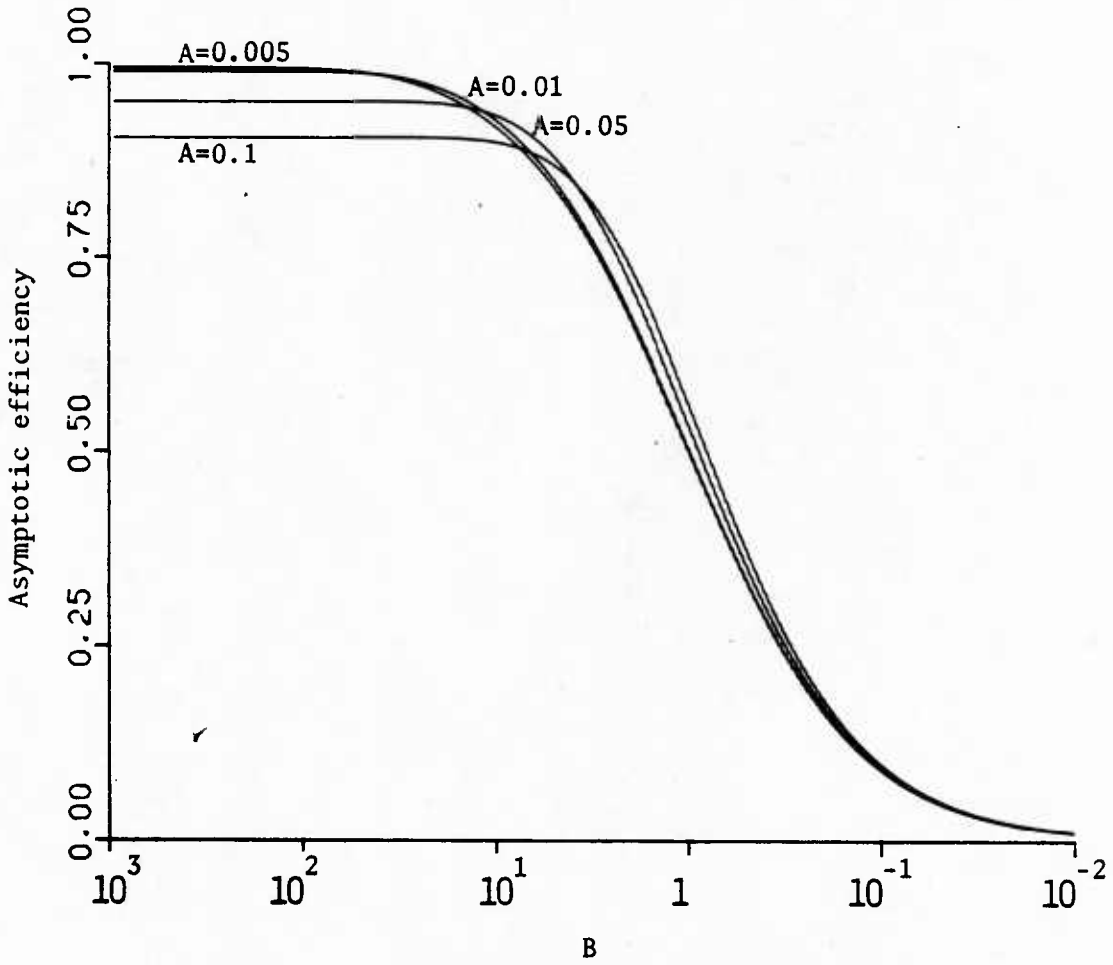

Figure 10 Asymptotic efficiency of LS estimator vs. B for Middleton's class A process

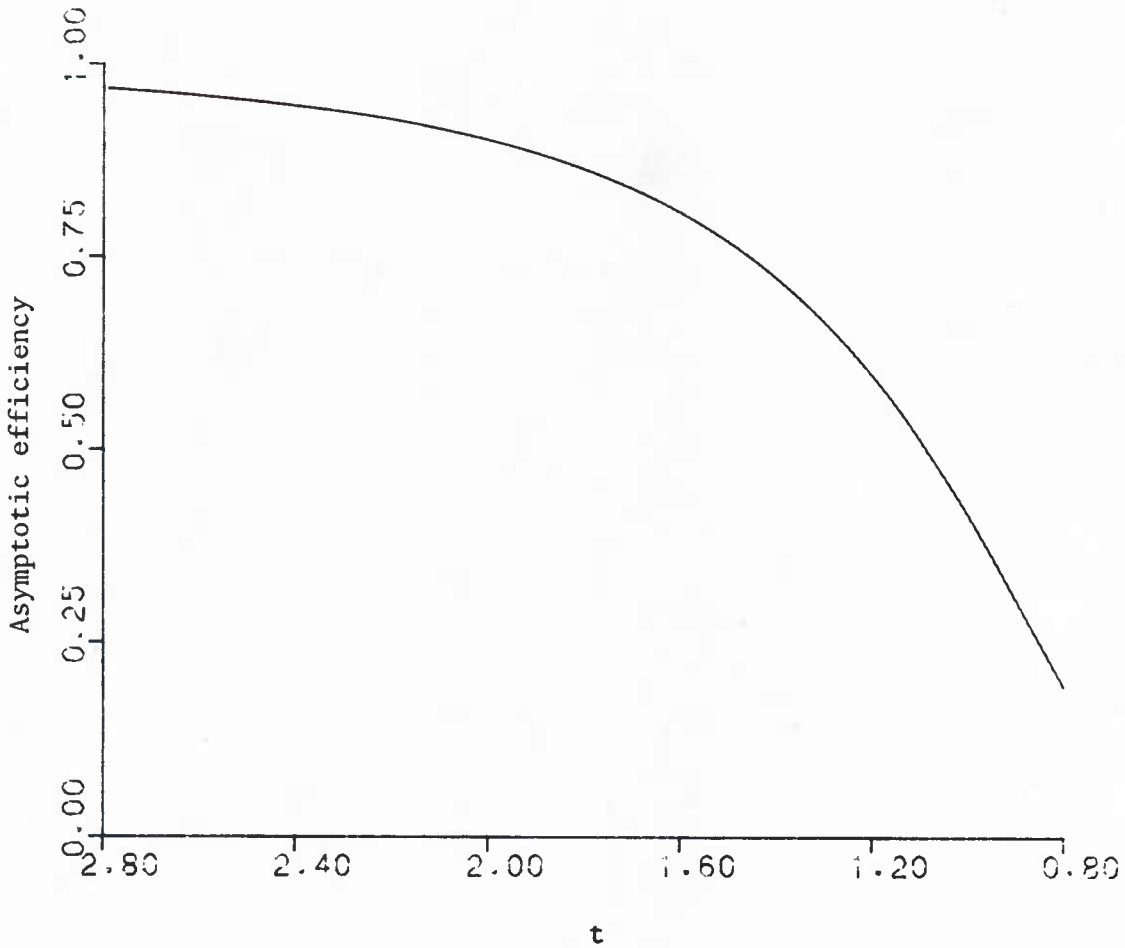

Figure 11 Asymptotic efficiency of LS estimator vs. t for Johnson's process

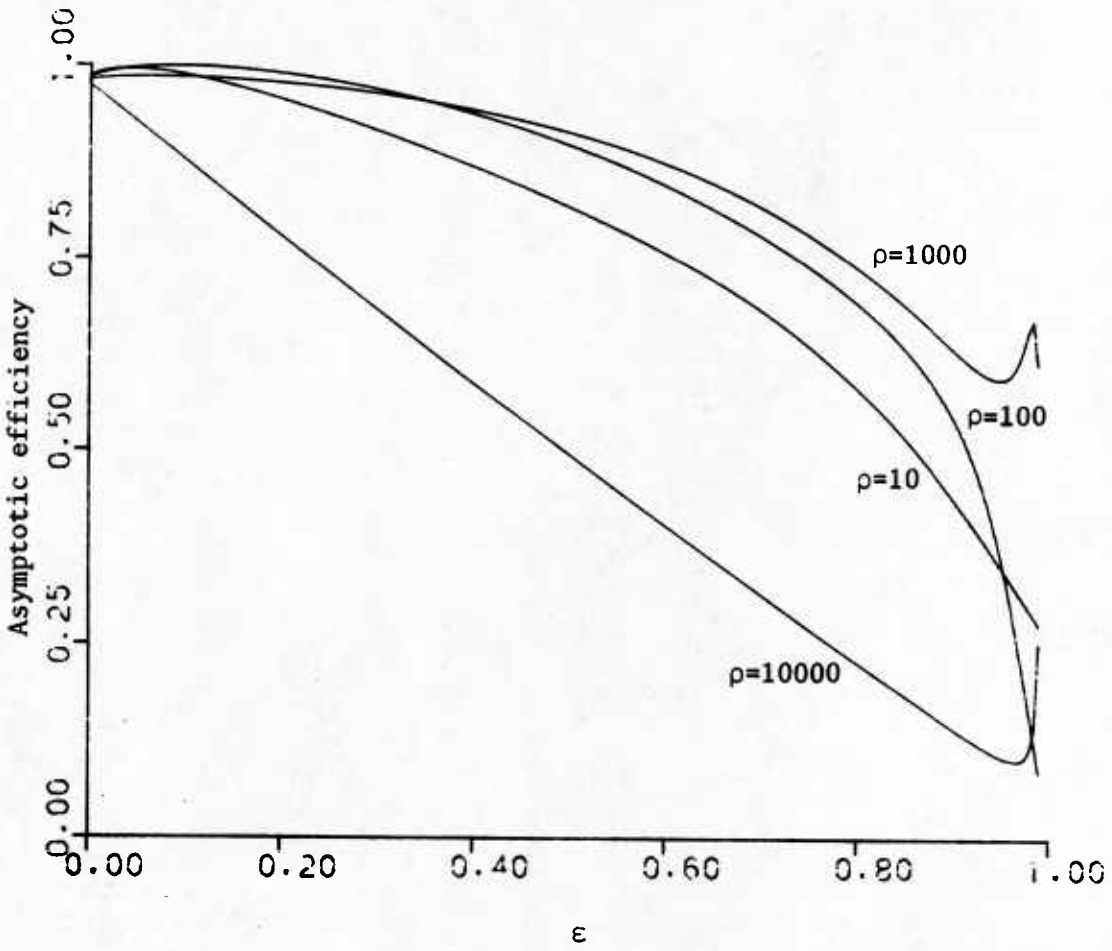

Figure 12 Asymptotic efficiency of M-estimator vs. ε for the Mixed-Gaussian process

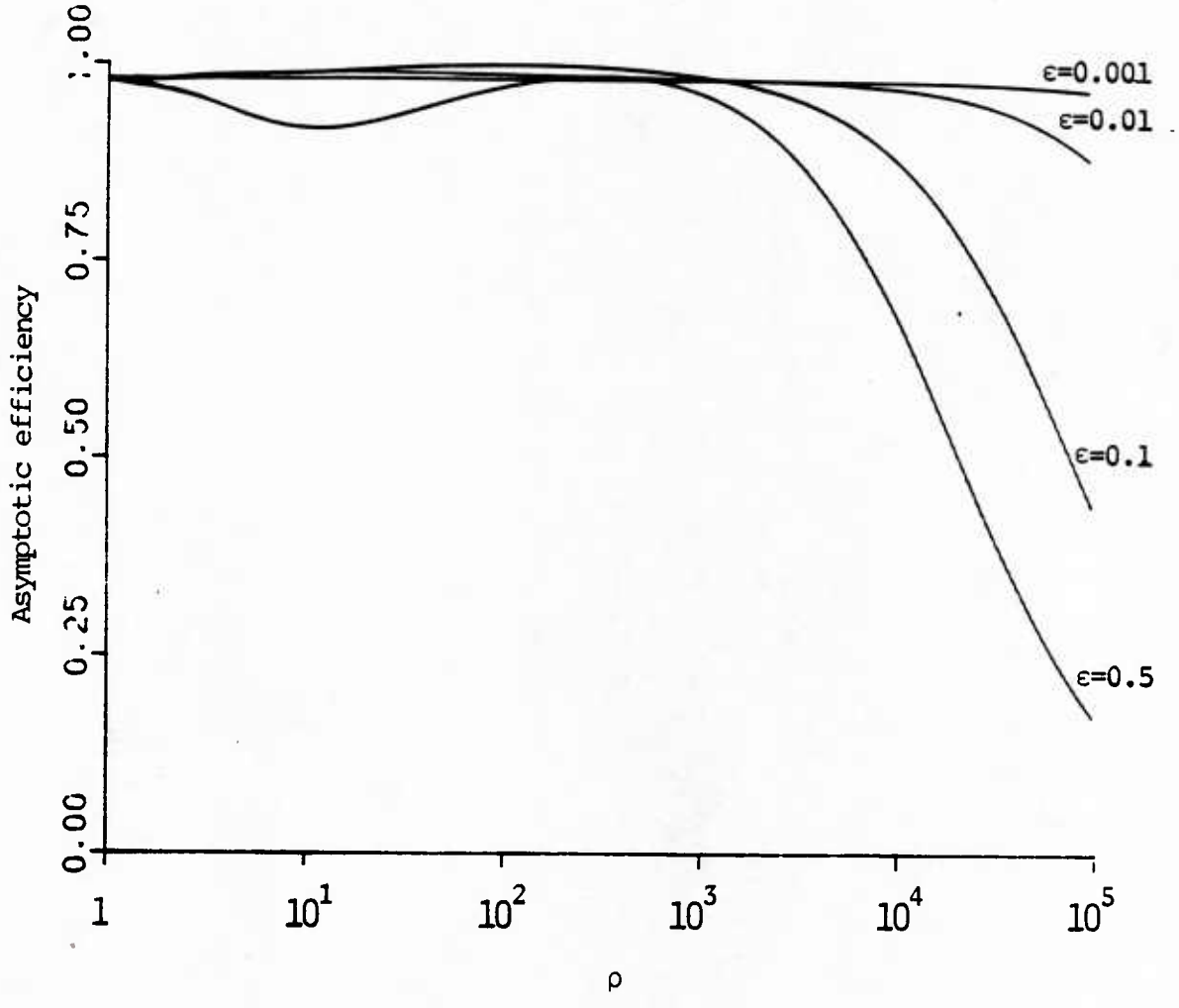

Figure 13 Asymptotic efficiency of M-estimator vs. ρ for the Mixed-Gaussian process

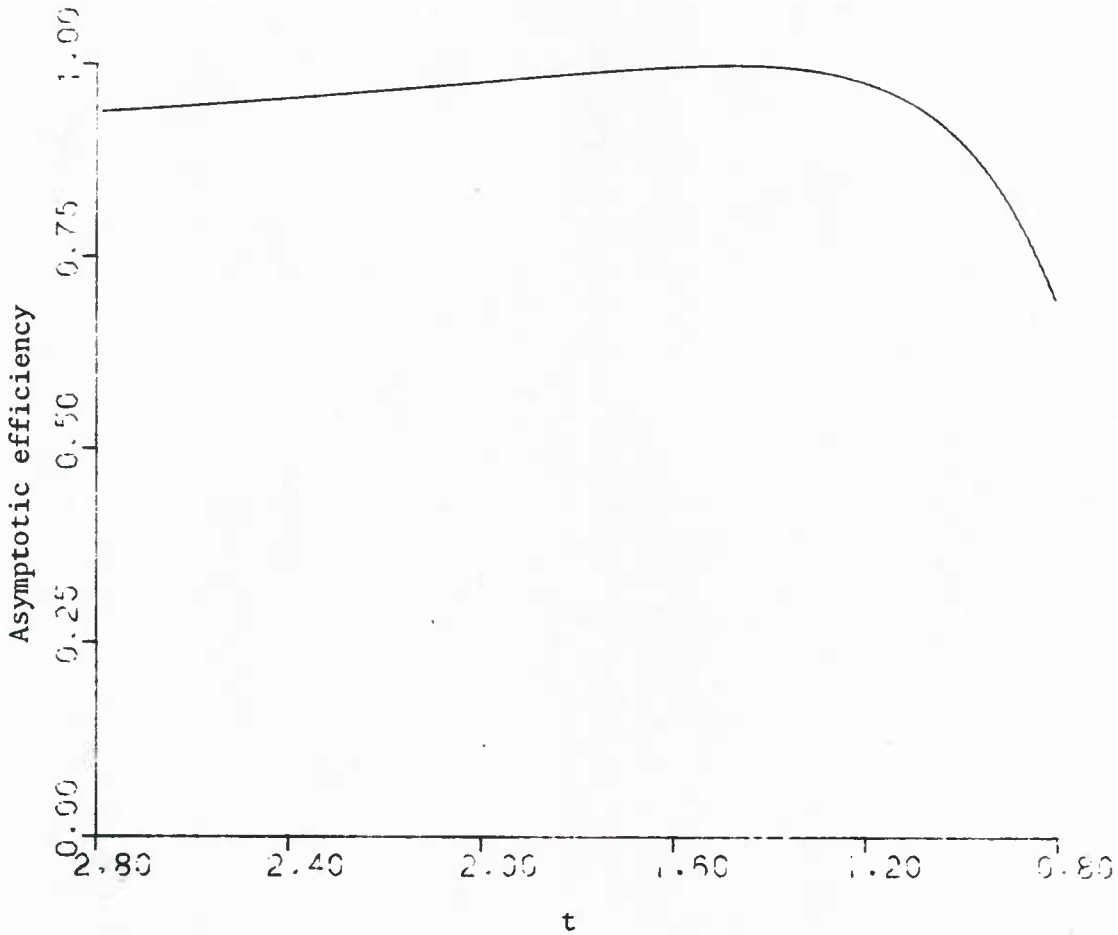

Figure 15 Asymptotic efficiency of M-estimator vs. t for Johnson's process

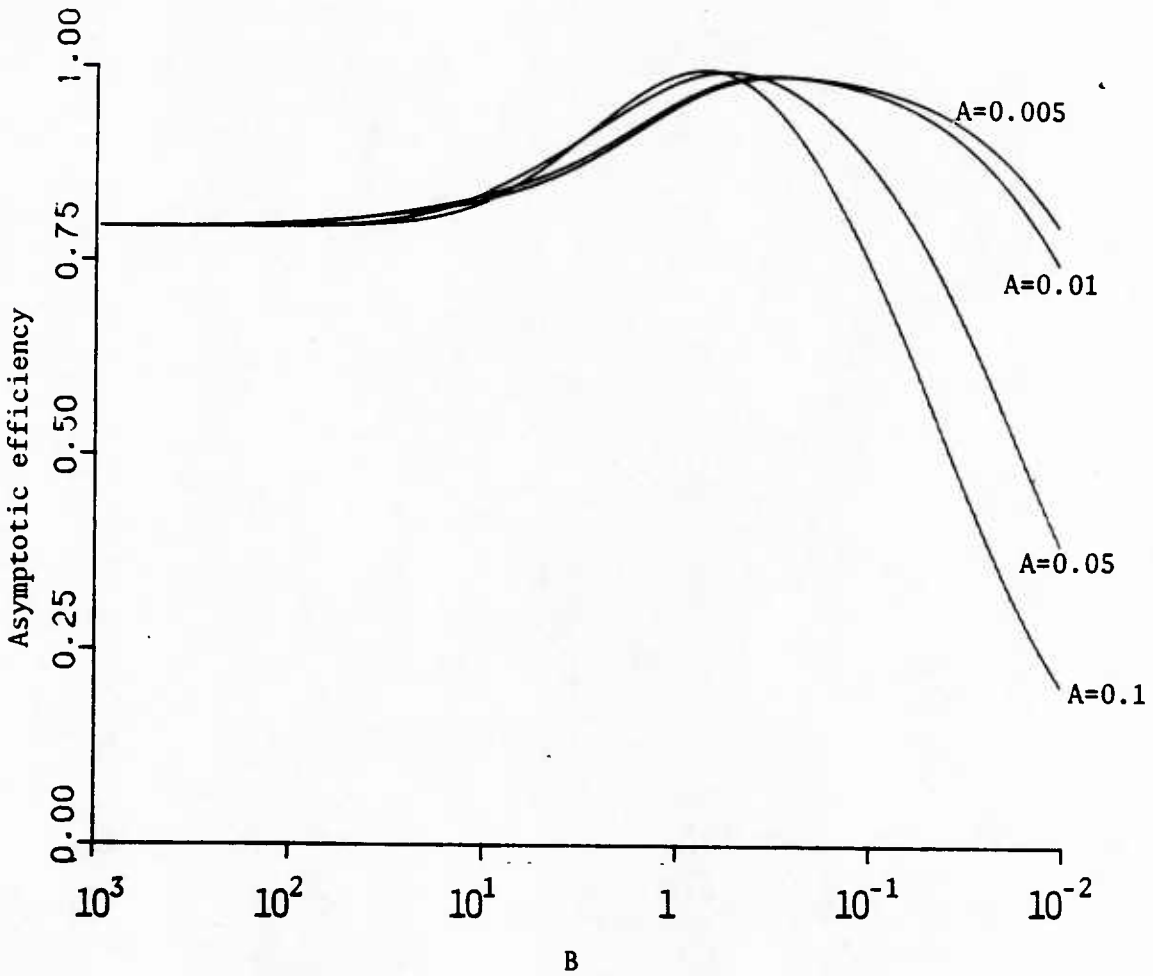

Figure 14 Asymptotic efficiency of M-estimator vs. B for Middleton's class A process

Simple and Efficient Estimation of Parameters of Non-Gaussian Autoregressive Processes

STEVEN KAY AND DEBASIS SENGUPTA
Department of Electrical Engineering
University of Rhode Island
Kingston, Rhode Island 02881

Abstract

A new technique for the estimation of autoregressive filter parameters of a non-Gaussian autoregressive process is proposed. The probability density function of the driving noise is assumed to be known. The new technique is a two-stage procedure motivated by maximum likelihood estimation. It is computationally much simpler than the maximum likelihood estimator and does not suffer from convergence problems. Computer simulations indicate that unlike the least squares or linear prediction estimators, the proposed estimator is nearly efficient, even for moderately sized data records. By a slight modification the proposed estimator can also be used in the case when the parameters of the driving noise probability density function are not known.

This work was supported by the Office of Naval Research under contract No. N00014-84-K-0527.

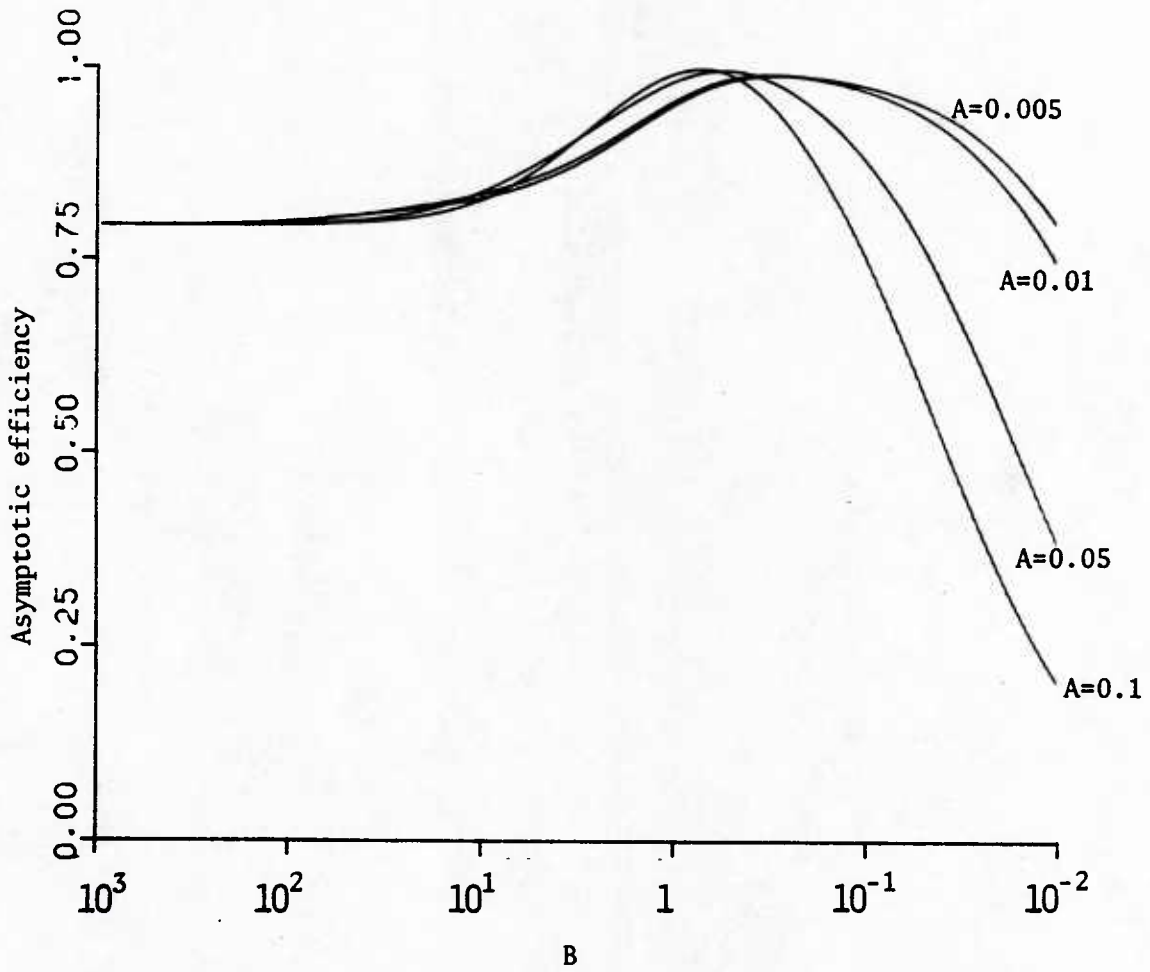

Figure 14 Asymptotic efficiency of M-estimator vs. B for Middleton's class A process

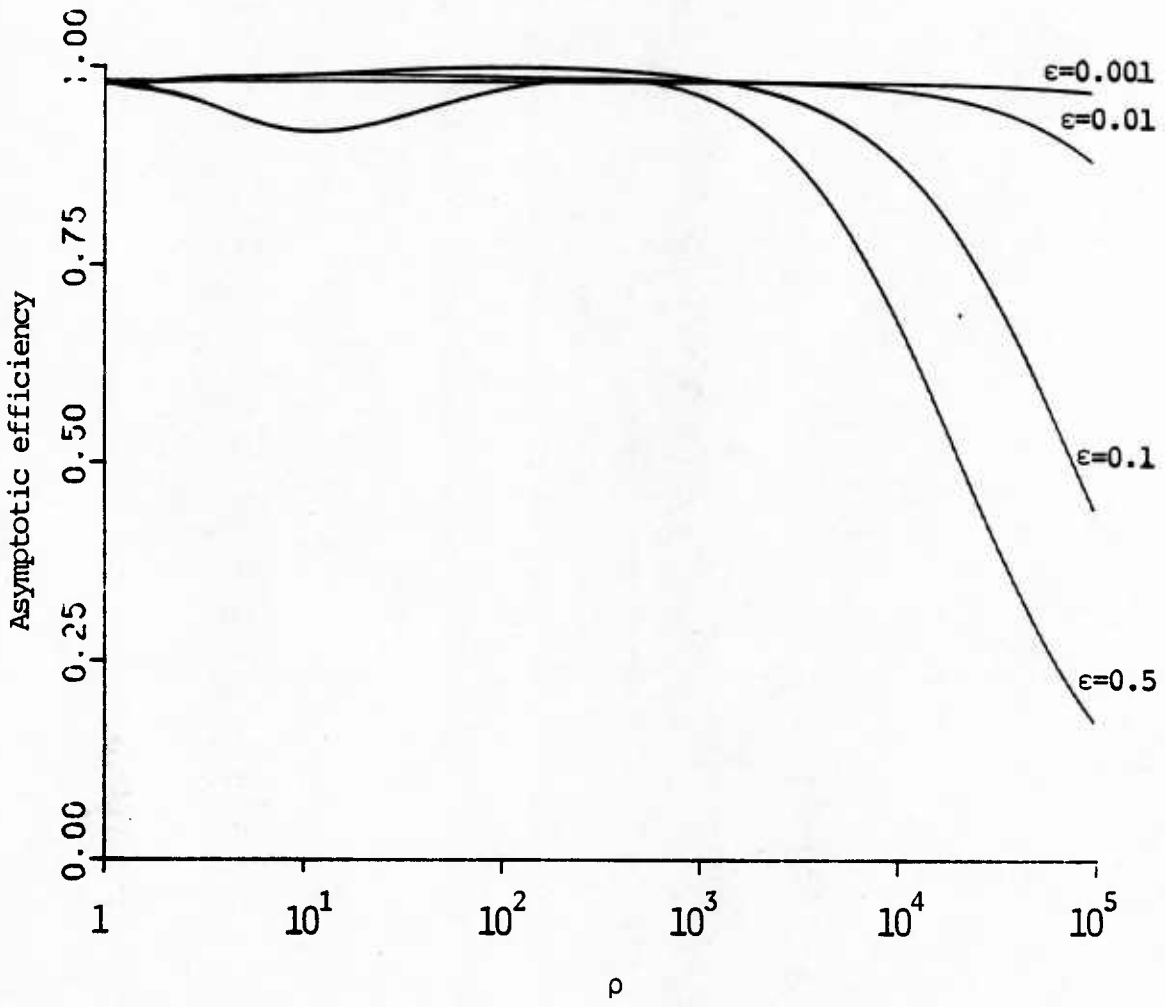

Figure 13 Asymptotic efficiency of M-estimator vs. ρ for the Mixed-Gaussian process

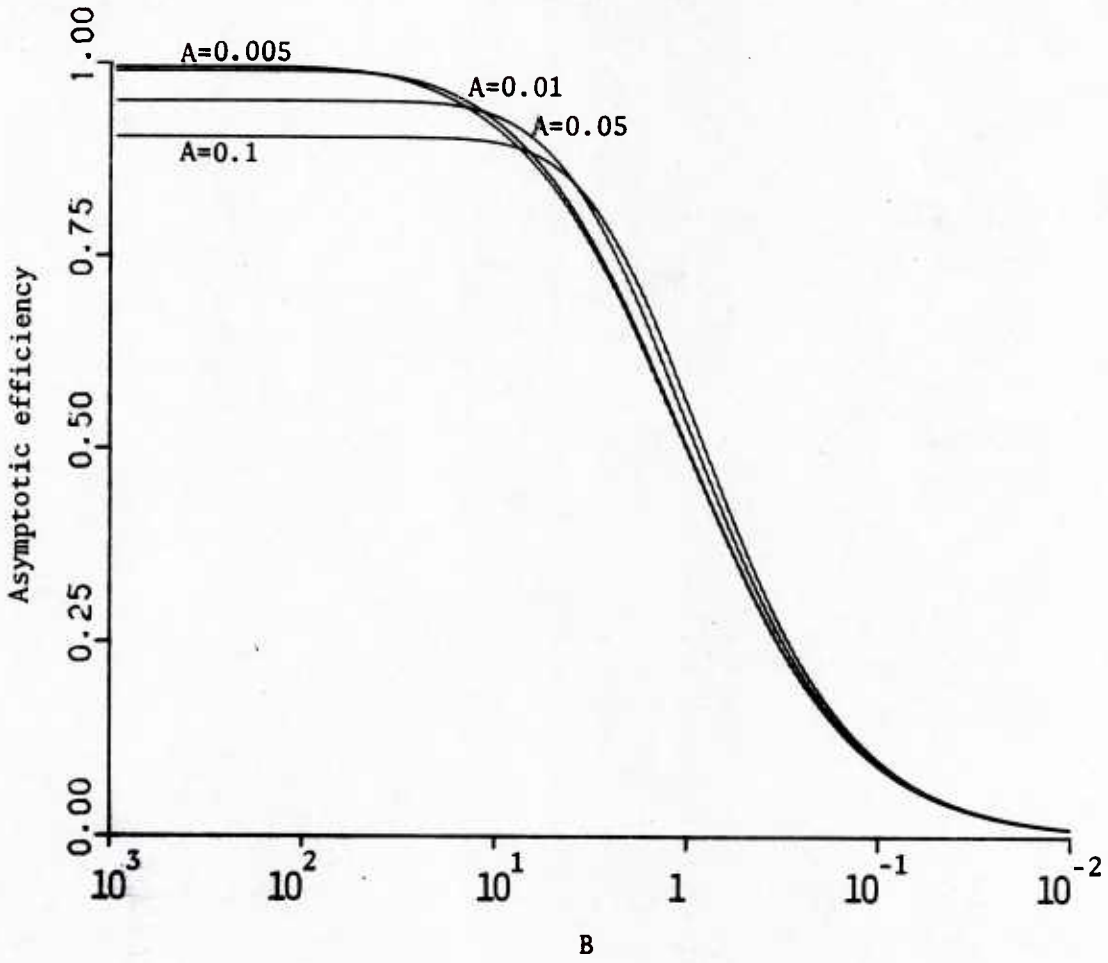

Figure 10 Asymptotic efficiency of LS estimator vs. B for Middleton's class A process

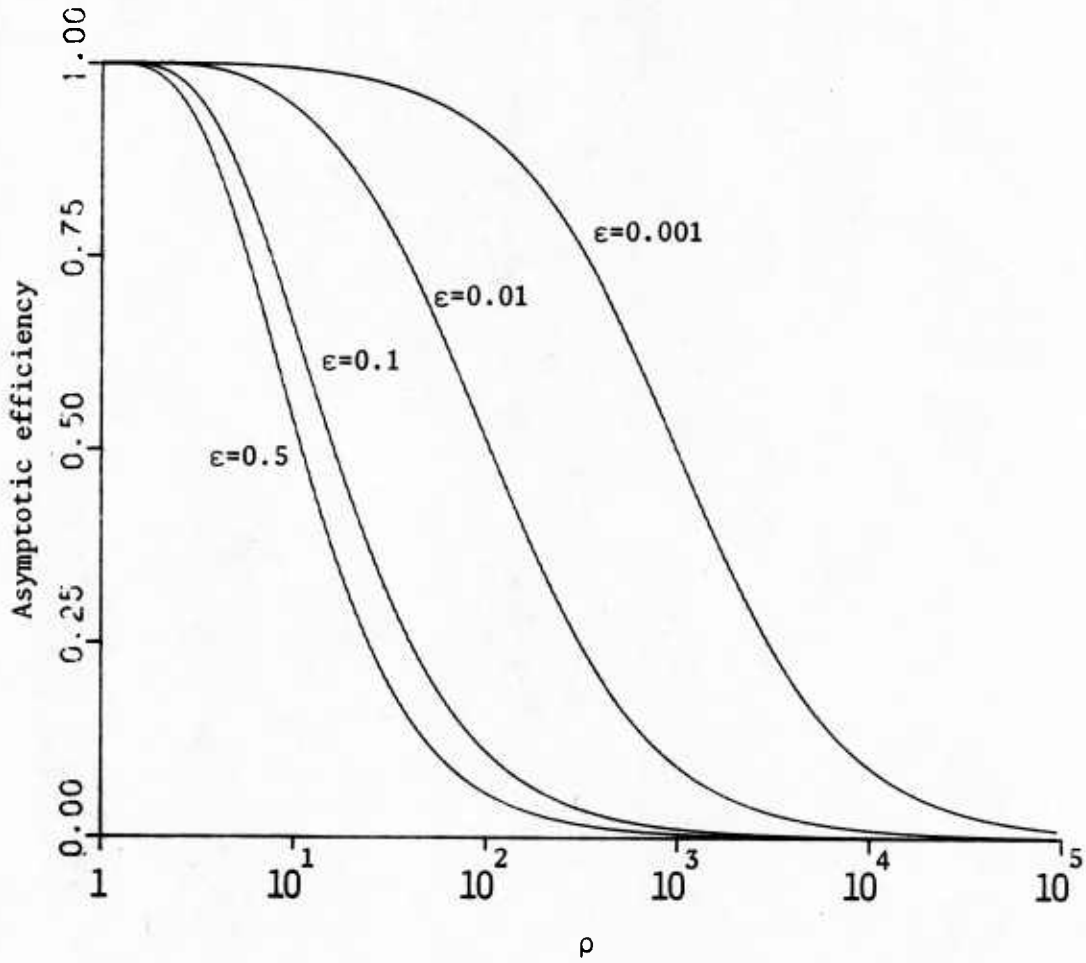

Figure 9 Asymptotic efficiency of LS estimator vs. ρ for the Mixed-Gaussian process

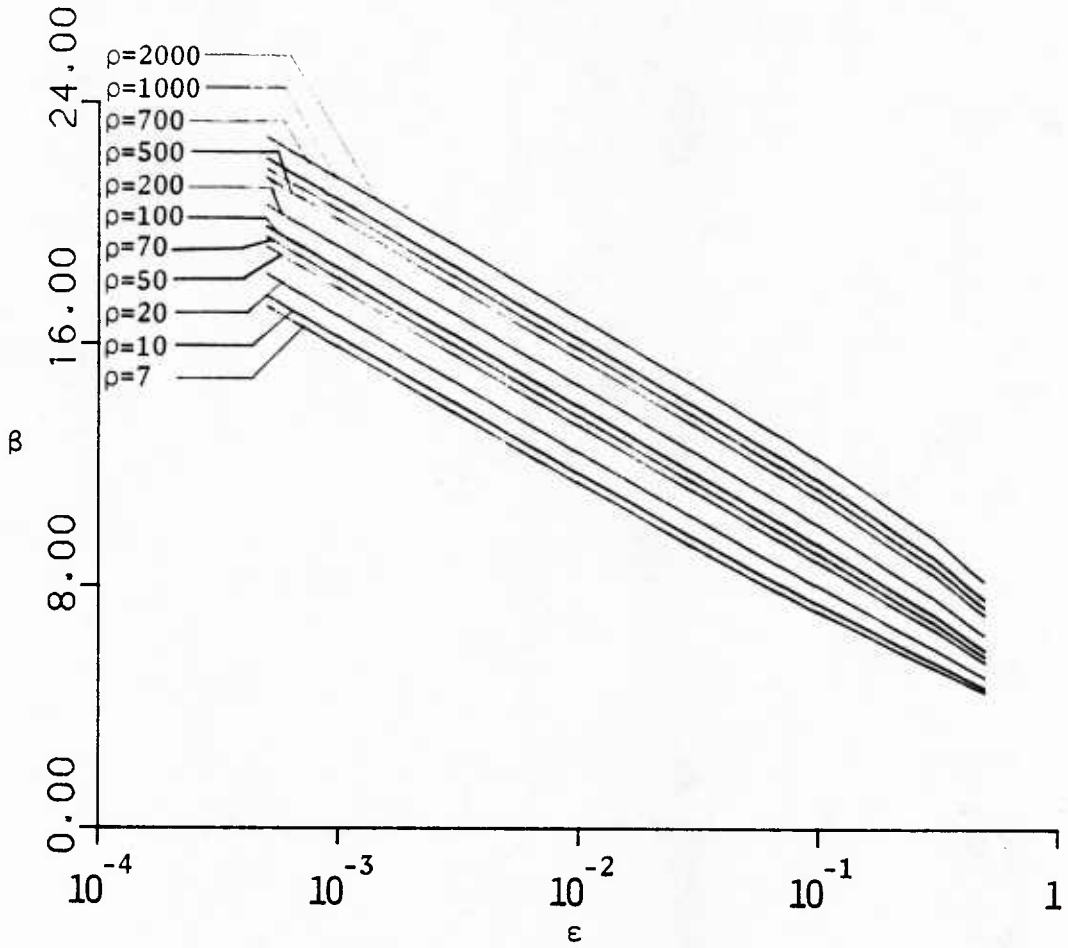

Figure 6 Dependence of β on ε and ρ in the Mixed-Gaussian case

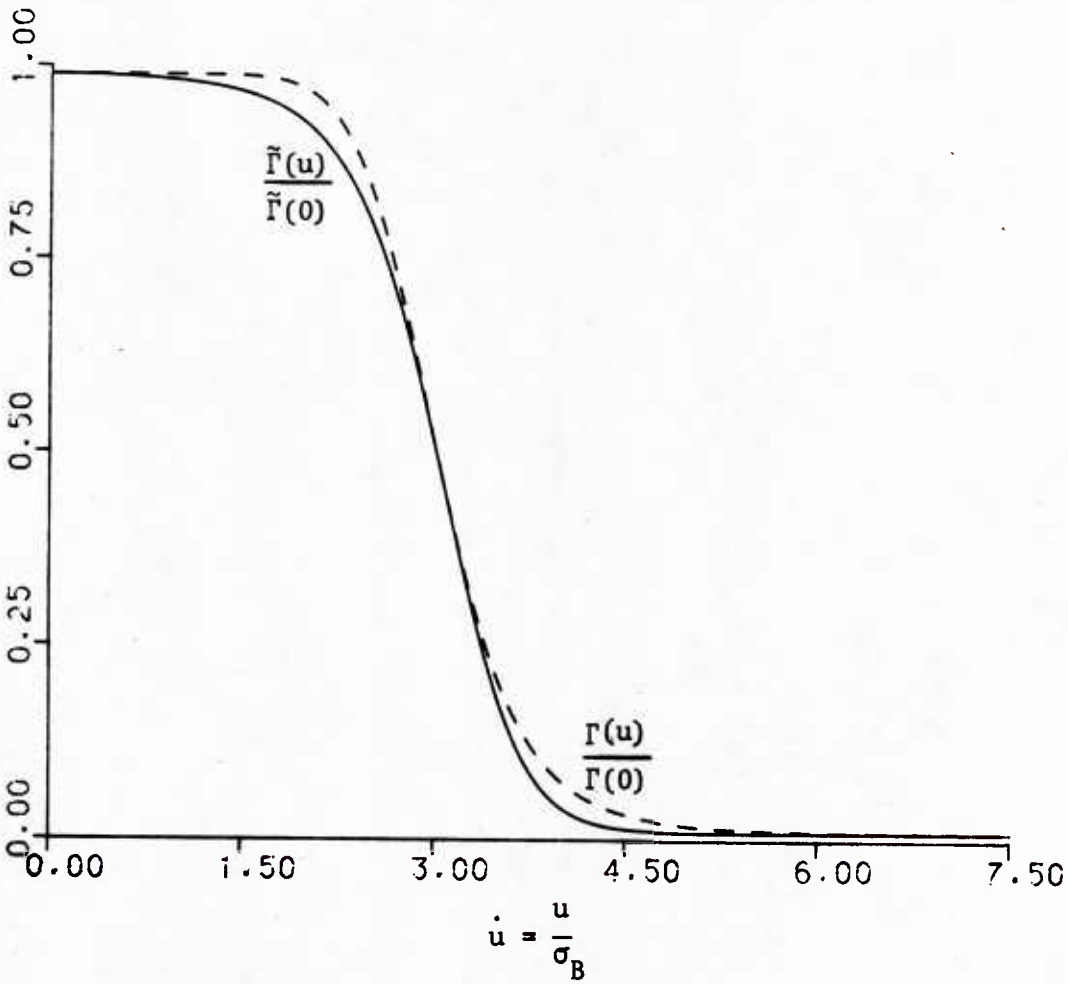

Figure 7 A typical approximation for the weighting curve in the Mixed-Gaussian case

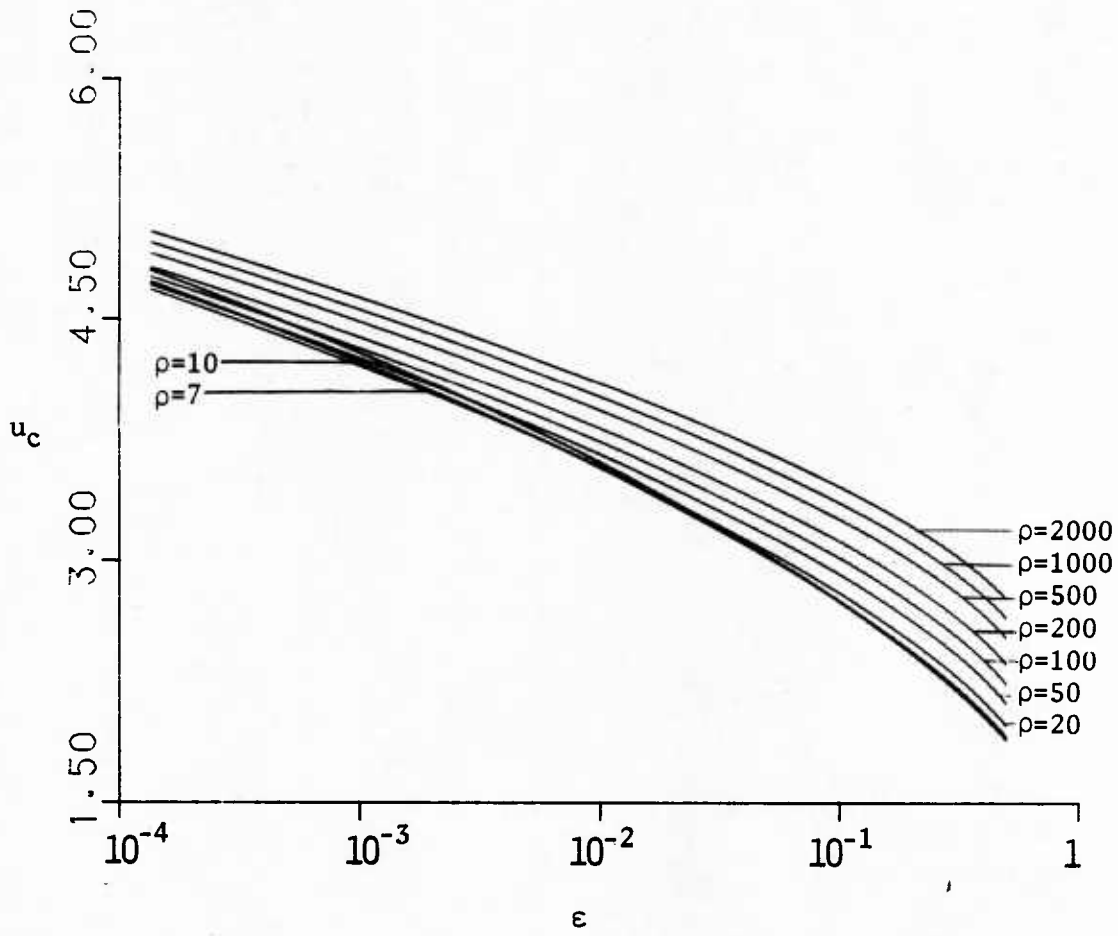

Figure 4 Dependence of threshold on ε in the Mixed-Gaussian case

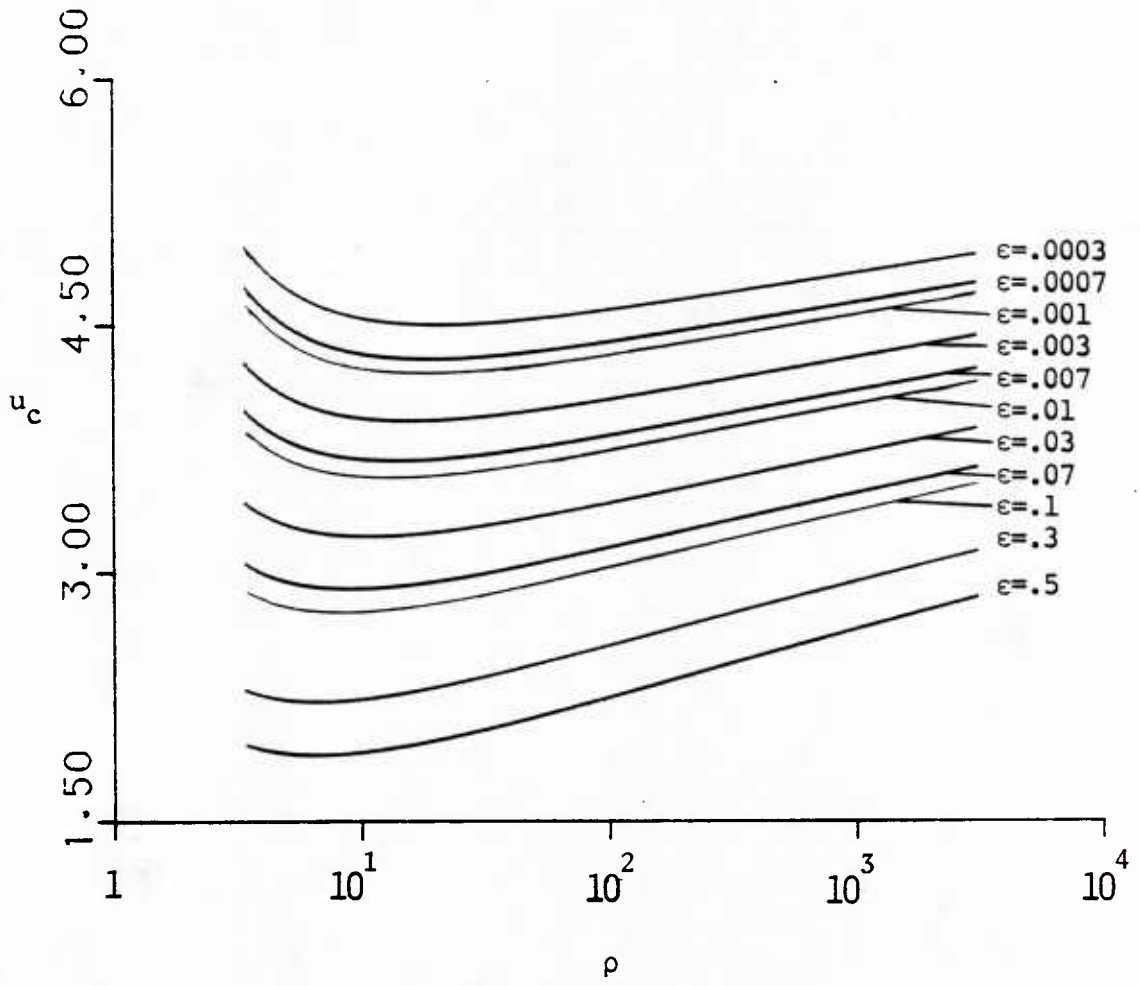

Figure 5 Dependence of threshold on ρ
in the Mixed-Gaussian case

OFFICE OF NAVAL RESEARCH
STATISTICS AND PROBABILITY PROGRAM

BASIC DISTRIBUTION LIST
FOR
UNCLASSIFIED TECHNICAL REPORTS

FEBRUARY 1982

	Copies
Statistics and Probability Program (Code 411(SP)) Office of Naval Research Arlington, VA 22217	3
Defense Technical Information Center Cameron Station Alexandria, VA 22314	12
Commanding Officer Office of Naval Research Eastern/Central Regional Office Attn: Director for Science Barnes Building 495 Summer Street Boston, MA 02210	1
Commanding Officer Office of Naval Research Western Regional Office Attn: Dr. Richard Lau 1030 East Green Street Pasadena, CA 91101	1
U. S. ONR Liaison Office - Far East Attn: Scientific Director APO San Francisco 96503	1
Applied Mathematics Laboratory David Taylor Naval Ship Research and Development Center Attn: Mr. G. H. Gleissner Bethesda, Maryland 20084	1
Commandant of the Marine Coprs (Code AX) Attn: Dr. A. L. Slafkosky Scientific Advisor Washington, DC 20380	1
Navy Library National Space Technology Laboratory Attn: Navy Librarian Bay St. Louis, MS 39522	1
U. S. Army Research Office P.O. Box 12211 Attn: Dr. J. Chandra Research Triangle Park, NC 27706	1
Director National Security Agency Attn: R51, Dr. Maar Fort Meade, MD 20755	1
ATAA-SL, Library U.S. Army TRADOC Systems Analysis Activity Department of the Army White Sands Missile Range, NM 88002	1
ARI Field Unit-USAREUR Attn: Library c/o ODCSPER HQ USAEREUR & 7th Army APO New York 09403	1
Library, Code 1424 Naval Postgraduate School Monterey, CA 93940	1
Technical Information Division Naval Research Laboratory Washington, DC 20375	1
OASD (I&L), Pentagon Attn: Mr. Charles S. Smith Washington, DC 20301	1

	Copies
Director AMSAA Attn: DRXSY-MP, H. Cohen Aberdeen Proving Ground, MD 21005	1
Dr. Gerhard Heiche Naval Air Systems Command (NAIR 03) Jefferson Plaza No. 1 Arlington, VA 20360	1
Dr. Barbara Bailar Associate Director, Statistical Standards Bureau of Census Washington, DC 20233	1
Leon Slavin Naval Sea Systems Command (NSEA 05H) Crystal Mall #4, Rm. 129 Washington, DC 20036	1
B. E. Clark RR #2, Box 647-B Graham, NC 27253	1
Naval Underwater Systems Center Attn: Dr. Derrill J. Bordelon Code 601 Newport, Rhode Island 02840	1
Naval Coastal Systems Center Code 741 Attn: Mr. C. M. Bennett Panama City, FL 32401	1
Naval Electronic Systems Command (NELEX 612) Attn: John Schuster National Center No. 1 Arlington, VA 20360	1
Defense Logistics Studies Information Exchange Army Logistics Management Center Attn: Mr. J. Dowling Fort Lee, VA 23801	1

	Copies
Reliability Analysis Center (RAC) RADC/RBRAC Attn: I. L. Krulac Data Coordinator/ Government Programs Griffiss AFB, New York 13441	1
Technical Library Naval Ordnance Station Indian Head, MD 20640	1
Library Naval Ocean Systems Center San Diego, CA 92152	1
Technical Library Bureau of Naval Personnel Department of the Navy Washington, DC 20370	1
Mr. Dan Leonard Code 8105 Naval Ocean Systems Center San Diego, CA 92152	1
Dr. Alan F. Petty Code 7930 Naval Research Laboratory Washington, DC 20375	1
Dr. M. J. Fischer Defense Communications Agency Defense Communications Engineering Center 1860 Wiehle Avenue Reston, VA 22090	1
Mr. Jim Gates Code 9211 Fleet Material Support Office U. S. Navy Supply Center Mechanicsburg, PA 17055	1
Mr. Ted Tupper Code M-311C Military Sealift Command Department of the Navy Washington, DC 20390	1

Copies

Copies

Mr. F. R. Del Priori
Code 224
Operational Test and Evaluation
Force (OPTEVFOR)
Norfolk, VA 23511 1