BROAD SPECTRUM ESTIMATES AND CONFIDENCE INTERVALS FOR TAIL QUANTILES*

by

Leo Breiman

and

Charles J. Stone

Technical Report No. 46
July 1985

*Research Partially Supported by
Office of Naval Research Contract N00014-82-K-0054.

Department of Statistics
University of California
Berkeley, California

BROAD SPECTRUM ESTIMATES AND CONFIDENCE INTERVALS FOR TAIL QUANTILES*

by

Leo Breiman
and
Charles J. Stone

University of California, Berkeley

Abstract

A method is given for computing estimates and confidence bounds for extreme tail quantiles. The idea is to fit a two-parameter "quadratic tail model" to the upper tail of the data. Extensive simulation experiments show that the method gives accurate results over a wide range of distributions "centered at the exponential" which are neither too heavy- nor too light-tailed.

* Work supported by Office of Naval Research under contract N00014-82-K-0054.

Key words: Tail estimates, tail models.

1. Introduction. Often, in applications, one has a limited data set and wants to estimate some upper tail quantiles. For instance, one might have 30 years of annual high water levels on a river and want to estimate the 100 year flood level $x_{.01}$ defined by the requirement that the probability of an annual water level exceeding $x_{.01}$ is .01.

This paper discusses a method for estimating both tail quantiles and confidence intervals for these quantiles. Monte Carlo experiments show that the method produces coverage probabilities that are accurate for a wide range of distributions "centered at the exponential" which are neither too heavy- nor too light-tailed, for tail quantiles which are not too extreme, and for moderate to large sample sizes.

Although the estimation of tail parameters is becoming an increasingly more frequent practical problem, very little attention has been given to it in published statistical research. Because of this, we give some introductory remarks more lengthy than usual.

The statistical problem of estimating central parameters, i.e. location and spread, is analogous to interpolation of functions. That is, they are estimation problems "interior to the data". In contrast, estimating tail parameters is analogous to extrapolation or estimating parameters "exterior to the data".

The standard (and dangerous) method for approaching this tail estimation problem has been to assume that the data is sampled independently from a specified parametric family, say Weibull or lognormal. Then the parameters are estimated, usually by maximum likelihood, and the resulting distribution used to calculate the estimate for the upper tail quantile. The danger we refer to is that, particularly for heavy-tailed distributions, the estimates can differ drastically depending on the parametric form assumed for the fitted distribution.

To pursue the function fitting analogy, suppose there is a function f(x), some of whose values are known on [0,1], and we want to extrapolate f(x) into the region

$x > 1$. One method, corresponding to the statistical method outlined above, is to fit $f(x)$ on $[0,1]$ by a parametric family, i.e. a quadratic or cubic polynomial, and then use the fitted function to extrapolate. It is known in numerical analysis that the results can be disastrous.

A more accurate method is to extrapolate $f(x)$ by estimating its behavior in the vicinity of $x = 1$. For instance, one could use the known values of $f(x)$ near $x = 1$ to estimate $f(1), f'(1), f''(1)$ and then extrapolate using the 2^{nd} order Taylor expansion

$$f(x+1) \doteq f(1) + xf'(1) + \frac{x^2}{2} f''(1) .$$

If $f(x)$ is sufficiently smooth, this method will give accurate extrapolation over a limited range of $x > 1$.

An important aspect of the 2^{nd} order Taylor expansion and other neighborhood extrapolation methods is that ***they do not depend on the global behavior of*** $f(x)$ ***over*** $[0,1]$. The method we study is analogous to a 2^{nd} order Taylor expansion. It fits only the upper tail of the data. By ignoring the global behavior of the bulk of the data, it is thus applicable to a wide range of distributions. It is intermediate between the usual parametric and nonparametric models. It is parametric, in the sense that it is fitted to the tail of the data by estimating some parameters. It is nonparametric in the analogous sense that the applicability or accuracy of a Taylor expansion does not require that $f(x)$ be in a specified parametric family of functions, but only have a certain degree of smoothness around the extrapolation region. Because of these properties, we call our procedure "broad spectrum" in analogy with medical terminology for antibiotics.

Another point is important. As statisticians, we strongly believe that any estimate of a parameter should be accompanied by an indication of its accuracy. For this reason, the major thrust of this present work is to get accurate confidence intervals for tail quantiles. It may be discouraging to some practitioners to see how large the confidence intervals become for the more heavy-tailed distributions and extreme quantiles studied (see Section 4.5). But evidence is provided in Section 4.6 that this is in the nature of the problem and not a shortcoming of the method.

The layout of this paper is as follows: Section 2 discusses the tail-fitting models and their use in deriving estimates and confidence intervals for tail quantiles. Section 3 covers the "range of extrapolation". That is, the range of distributions, tail quantiles, and sample sizes to which our method is applicable. Section 4 gives the results of the Monte Carlo simulations, and Section 5 gives final remarks and conclusions.

Previous work on inference about the upper tail of a distribution has focussed on methods of estimation that are appropriate when the tail is (I) in the domain of attraction of some extreme value distribution; (II) approximately algebraically decreasing; or (III) approximately exponentially decreasing. These three conditions are very closely related. For example, the upper tail of X is approximately algebraically decreasing if and only if that of $\log X$ is approximately exponentially decreasing. In category (I) are Maritz and Munro (1967), Pickands (1975), Weissman (1978), Boos (1984), and Davis and Resnick (1984); in category (II) are Hill (1975), DuMouchel and Olshen (1975), DuMouchel (1983), Hall and Welsh (1985), and Csörgő, et al. (1985); and in category (III) are Breiman, et al. (1978, 1979 and 198l) and Crager (1982).

2. The Extrapolation Method.

2.1 *The exponential tail model.* The problem that began our research was an EPA funded project to find methods for estimating upper tail quantiles in air pollution (Breiman, Stone, and Gins, 1978). Typically, the sample size n was about 200 to 300, and the tail quantiles x_p of interest had p in the neighborhood of $1/n$.

At that time (1976-78) there were disputes among the experts as to whether daily air pollution levels could be best fit by Weibull, lognormal, or gamma distributions. The estimates of the upper tail quantiles differed considerably depending on the family of distributions hypothesized as truth.

The fitted distributions ranged from moderately heavy-tailed to slightly light-tailed. For instance, if Weibulls were fitted to a large variety of air pollution data sets, the power parameter varied from about .7 for the heavy-tailed to about 1.2 for the light-tailed.

On closer inspection, we saw that the tails of the various distributions were similar in shape and could be reasonably approximated by an exponential tail fit of the form

$$G(x) \doteq p_0 e^{-(x-x_{p_0})/a}, \quad x \geq x_{p_0},$$

where $F(x)$ is the d.f., $G(x) = 1 - F(x)$ and $x_p = G^{-1}(p)$. This leads to the approximation

$$x_p \doteq x_{p_0} + a \log(p_0/p), \quad p \leq p_0$$

and suggested the following estimation strategy: take p_0 in the range .1 - .3; set $m = [np_0]$, estimate x_{p_0} by $x_{(m)}$, the mth order statistic of the sample $x_1, \ldots, x_n$; estimate a by

$$\hat{a} = \frac{1}{m-1} \sum_{i}^{m} (X_{(i)} - X_{(m)})$$

and x_p by

$$\hat{x}_p = x_{(m)} + \hat{a} \log[m/(n+1)p] .$$

The initial 1978 study and more detailed subsequent studies (Breiman and Stone, 1979; Breiman and Stone, 1981) showed that the tail exponential fit gives tail quantile estimates that are surprisingly accurate in a mean-squared sense over a wide spectrum of distributions. This was particularly gratifying since the sample sizes used $(n = 100$ to $1000)$ were nowhere near the requirements for the asymptotic exponentiality of the tails.

The next step in our investigation was an effort to construct confidence bounds for tail quantiles. Here, we discovered that exponential tail estimates were unsuitable for confidence interval construction. The reason is interesting. In estimating central parameters such as location or spread, an estimate with low mean squared error usually has low bias, and most of the mean squared error is variance. In estimating tail parameters, a major source of error is in the occasional large overshot. A good mean squared error estimate will tend to have a substantial downward bias in order to reduce the overshooting. And, in fact, for long-tailed distributions, the exponential tail estimate's good mean squared error performance went hand-in-hand with a marked downward bias.

2.2 ***The quadratic tail model.*** In the exponential tail model x_p is approximated by an expression that in linear in $\log p$. To get a more nearly unbiased estimate we use the tail model

$$x_p \doteq x_{p_0} + a \log(p_0/p) - \frac{b}{2}(\log^2 p_0 - \log^2 p), \quad 0 < p \leq p_0 , \tag{2.1}$$

which is quadratic in $\log p$. For the exponential tail model,

$$x_p \doteq x_{p_0} + a \log(p_0/p), \quad 0 < p \leq p_0$$

it was natural to estimate x_{p_0} by $X_{[np_0]}$ and a by a linear combination of order statistics $X_{(i)}$ with $i \geq [np_0]$. For the quadratic tail model we adopted the same strategy, estimating x_{p_0} by $X_{(m)}$, $m = [np_0]$, and a, b by linear combinations of order statistics $X_{(i)}$, $i \geq m$. Of course, this leads to the question of determining the coefficients of these linear combinations. Our approach to this is to assume that (2.1) holds exactly, and to determine coefficients such that the estimates $\hat{a}, \hat{b}$ are "unbiased and minimum variance" in a sense made more precise below. To do this, means and variances need to be computed, so that the first order of business is to establish the distributions of the order statistics $X_{(i)}$, $i \geq m$, under the tail model (2.1).

In (2.1) put $p = e^{-y}$, getting

$$G^{-1}(e^{-y}) \doteq x_{p_0} + a \log p_0 + y - \frac{b}{2}(\log^2 p_0 - y^2)\,, \tag{2.2}$$

for $e^{-y} \leq p_0$. Let $U_{(1)} \geq \dots \geq U_{(n)}$ be the order statistics based on n independent uniforms on [0,1]. Then $G(X_{(1)}), \dots, G(X_{(n)})$ have the same joint distribution as $1 - U_{(1)}, \dots, 1 - U_{(n)}$. Let $Z_1, \dots, Z_n$ be n independent random variables, each having an exponential distribution with mean 1, and $Z_{(1)} \geq \dots \geq Z_{(n)}$ the corresponding order statistics. Then $1 - U_{(1)}, \dots, 1 - U_{(n)}$ have the same joint distribution as $e^{-Z_{(1)}}, \dots, e^{-Z_{(n)}}$, and $X_{(1)}, \dots, X_{(n)}$ have the same joint distribution as $G^{-1}(e^{-Z_{(1)}}), \dots, G^{-1}(e^{-Z_{(n)}})$. Therefore, by (2.2) the joint distribution of $X_{(1)}, \dots, X_{(m)}$ is approximately the same as that of the variables

$$x_{p_0} + a(\log p_0 + Z_{(k)}) - \frac{b}{2}(\log^2 p_0 - Z_{(k)}^2), \quad k = 1, \dots, m. \tag{2.3}$$

In particular, the variables $X_{(k)} - X_{(k+1)}$, $1 \leq k \leq m - 1$ have the same joint distribution as

$$a(Z_{(k)} - Z_{(k+1)}) + \frac{b}{2}(Z_{(k)}^2 - Z_{(k+1)}^2), \quad k = 1, \dots, m-1\,. \tag{2.4}$$

Working with these variables is made tractable by using the property that $Z_{(1)}, \ldots, Z_{(n)}$ have the same joint distribution as the variables

$$\sum_{k}^{n} \frac{Z_j}{j}, \quad k = 1, \ldots, n . \tag{2.5}$$

Now for any integer m, and $p_1 = m/(n+1) \leq p_0$, and $p \leq p_1$, (2.1) can be written as

$$x_p = x_{p_1} + a(\log p_1 - \log p) - \frac{b}{2}(\log^2 p_1 - \log^2 p) . \tag{2.6}$$

Then x_{p_1} can be estimated by $X_{(m)}$. To estimate the other terms in (2.6), note that p_1, p are known, so we are trying to form estimates of parameters of the form

$$r = La + Mb ,$$

with known L, M. Let $w_1, \ldots, w_{m-1}$ be constants, set

$$\hat{r} = \sum_{1}^{m-1} k w_k (X_{(k)} - X_{(k+1)}) ,$$

and consider $\hat{r}$ to be an estimate of r. We will call $\hat{r}$ unbiased if

$$E\hat{r} = La + Mb$$

for all values of a, b.

To see what conditions this imposes on the w_k, $k = 1, \ldots, m-1$, note that by (2.4) $\hat{r}$ has the same distribution as

$$\sum_{1}^{m-1} k w_k [a(Z_{(k)} - Z_{(k+1)}) + \frac{b}{2}(Z_{(k)}^2 - Z_{(k+1)}^2)] . \tag{2.7}$$

By (2.5)

$$E(Z_{(k)} - Z_{(k+1)}) = E\frac{Z_k}{k} = \frac{1}{k} ;$$

so the mean of the first term in (2.7) equals $a \sum_{1}^{m-1} w_k$. In the second term, write

$$Z_{(k)}^2 - Z_{(k+1)}^2 = 2Z_{(k+1)}(Z_{(k)} - Z_{(k+1)}) + (Z_{(k)} - Z_{(k+1)})^2$$

and use (2.5) again to get

$$E(Z_{(k)}^2 - Z_{(k+1)}^2) = \frac{2u_k}{k}$$

where

$$u_k = \sum_{k}^{a} \frac{1}{j} .$$

Therefore,

$$E\hat{r} = a \sum_{1}^{m-1} w_k + b \sum_{1}^{m-1} u_k w_k$$

and unbiasedness requires that

$$L = \sum_{1}^{m-1} w_k , \quad M = \sum_{1}^{m-1} u_k w_k . \tag{2.8}$$

The variance of $\hat{r}$ is derived by a simple but lengthy computation given in the Appendix. To summarize, let

$$u_k^{(2)} = \sum_{k}^{a} \frac{1}{j^2} ,$$

$$\overline{w}_k = \frac{1}{k} \sum_{1}^{k} w_j .$$

Then

$$\text{Var}(\hat{r}) = \sum_{1}^{m-1} (aw_k + b\overline{w}_k + bu_k w_k)^2 + b^2[\sum_{1}^{m-1} (u_k^{(2)} w_k^2) + (m-1)^2 u_{m-1}^{(2)} \overline{w}_{m-1}^2] . \tag{2.9}$$

We would like to keep the variance as small as possible subject to (2.8), but no permissable choice of w_k will minimize (2.9) for all a, b. Instead, we notice that for b close to zero, $\mathrm{Var}(\hat{r}) \doteq a^2 \sum_1^{m-1} w_k^2$. So that for tail distributions not too far from exponential, a reasonable procedure is to select weights w_k to minimize $\sum_1^{m-1} w_k^2$ subject to the constraints (2.8).

It is geometrically clear that there is a unique solution to this minimization problem, given by $w_k = \lambda_1 + \lambda_2 u_k$ with λ_1, λ_2 selected to satisfy (2.8). The evaluation of λ_1, λ_2 is easy and results in

$$\lambda_1 = \frac{S_2 L - S_1 M}{D}, \quad \lambda_2 = \frac{(m-1)M - S_1 L}{D}$$

where $S_i = \sum_1^{m-1} u_k^i$, $i = 1,2$, and $D = (m-1)S_2 - S_1^2$. Therefore

$$\begin{aligned} w_k &= \frac{S_2 - S_1 u_k}{D} L + \frac{(m-1)u_k - S_1}{D} M \\ &= w_{1k} L + w_{2k} M . \end{aligned}$$

In particular, then, we take as our estimates of a and b,

$$\hat{a} = \sum_1^{m-1} k w_{1k} (X_{(k)} - X_{(k+1)})$$

$$\hat{b} = \sum_1^{m-1} k w_{2k} (X_{(k)} - X_{(k+1)})$$

and the estimate $\hat{r}$ of $La + Mb$ as $L\hat{a} + M\hat{b}$. The variance of $\hat{r}$ is given by (2.9) with $w_k = Lw_{1k} + Mw_{2k}$.

Thus, given a sample of size n from any distribution and given m as some fraction of n, the above procedure leads to estimates of any quantile x_p, $p < m/(n+1)$, based on the quadratic tail approximation. The Monte Carlo study in Section 4 shows that these estimates generally have low bias over the ranges of distributions and quantiles investigated.

2.3 *Setting confidence limits.* In (2.6), we set $p_1 = m/(n+1)$,

$$L = \log p_1 - \log p ,$$

and

$$M = -\frac{1}{2}(\log^2 p_1 - \log^2 p).$$

Then our estimate of x_p, $p < p_1$, is

$$\hat{x}_p = X_{(m)} + \hat{r}$$

where

$$\hat{r} = L\hat{a} + M\hat{b} .$$

To find confidence intervals for x_p, our first step is standard; find an estimate for $\mathrm{Var}(\hat{x}_p)$. Now

$$\mathrm{Var}(\hat{x}_p) = \mathrm{Var}(X_{(m)}) + 2\,\mathrm{Cov}(X_{(m)},\hat{r}) + \mathrm{Var}(\hat{r}) . \tag{2.10}$$

The variance of $\hat{r}$ is given by (2.9) as a quadratic expression in a and b. Similarly, the other two terms in (2.10) are quadratics in a and b; for example, from (2.3),

$$\mathrm{Var}(X_{(m)}) = \mathrm{Var}(aZ_{(m)} + \frac{b}{2}Z_{(m)}^2) .$$

Thus

$$\begin{aligned}\mathrm{Var}(\hat{x}_p) &= \sigma^2(a,b)\\ &= c_1a^2 + c_2ab + c_3b^2\end{aligned}$$

where c_1, c_2, c_3 depend only on n, m, L and M. The terms on the right in (2.10) are evaluated in the Appendix. As usual, we use $\sigma^2(\hat{a},\hat{b})$ as an estimate for the variance of $\hat{\tau}_p$.

Let Φ be the standard normal distribution function. Given α, $0 < \alpha < 1$, define z_α by $\Phi(z_\alpha) = 1-\alpha$. The usual large sample upper and lower $1-\alpha$ confidence founds for x_p are $x_p \pm z_\alpha \sigma(\hat{a},\hat{b})$. However, early in our Monte Carlo explorations we found that for tail quantiles the coverage given by these bounds may not be close to $1-\alpha$ unless m and n are surprisingly large. Even for exponential distributions which are exactly fit by a quadratic tail model, this lack of accuracy persists until very large sample sizes are used.

The main reason for this discrepancy is the positive correlation between $\hat{x}_p$ and $\sigma(\hat{a},\hat{b})$. This positive correlation results in coverage probabilities of the upper bound $\hat{x}_p + z_\alpha \sigma(\hat{a},\hat{b})$ that are significantly less than $1-\alpha$ and coverage probabilities for the lower bound significantly larger than $1-\alpha$.

Other contributing reasons are the unreliability of $\sigma^2(\hat{a},\hat{b})$ as an estimate of $\mathrm{Var}(\hat{x}_p)$ and, for moderate m, the error in Φ as an approximation to

$$\mathcal{L}\left(\frac{\hat{x}_p - x_p}{\sigma(a,b)}\right).$$

While we suspect (but haven't bothered to prove) that the usual large sample confidence bounds give asymptotically correct coverage, another approach is necessary to deal with realistic sample sizes.

After considerable experimenting, the following simple and effective remedy was found; use the upper and lower bounds

$$\hat{x}_p + \bar{t}_\alpha \sigma(\hat{a},\hat{b})$$

$$\hat{x}_p + \underline{t}_\alpha \sigma(\hat{a},\hat{b})$$

where $\overline{t}_\alpha$ is the upper α quantile of the distribution of $(x_p - \hat{x}_p)/\sigma(\hat{a},\hat{b})$ when the underlying variables $X_1, \ldots, X_n$ are independent, mean 1 exponentials, and $\underline{t}_\alpha$ is the lower α quantile of the same distribution.

While these two values are, in principle, analytically computable, as functions of m, n, p, and α, we have not been able to produce tractable expressions for them. Instead, we used a random number generator to repeatedly (N times) produce a sample of size n from the exponential distribution, fitted the quadratic tail model to each sample at $m/(n+1)$, and computed the resulting value of $(x_p - \hat{x}_p)/\sigma(\hat{a},\hat{b})$. Then $\overline{t}_\alpha, \underline{t}_\alpha$ were taken as the upper and lower α sample quantiles of these N values. We considered constructing a table of $\overline{t}_\alpha, \underline{t}_\alpha$ for various values of m, n, p, α but discarded it on the ground that anyone who programmed the quadratic tail procedure could easily derive the $\overline{t}_\alpha, \underline{t}_\alpha$ values desired by a procedure similar to the one we used.

3. The Range of Extrapolation.

3.1 ***What determines the accuracy?*** The range for which the method outlined above gives accurate results depends on the underlying distribution, the sample size, and how extreme the tail quantile to be estimated is. We have performed extensive simulation to determine the range of accuracy, and an abbreviated summary of results is given in Section 4. In this section we describe our basic set up for the experiment.

If one considers the exponential tail fit as a "tangent" approximation and the quadratic as adding a correction for "curvature", then it is sensible that the range of distributions for which our method is accurate center at the exponential and do not depart from it too drastically.

The Monte Carlo experiment described in Section 4 uses the Weibull, lognormal and generalized gamma distributions with choices of parameters that make them range from moderately heavy-tailed to moderately light-tailed. More explicitly, for the Weibull the most heavy-tailed distribution has a shape parameter of .52 and the most light-tailed a shape parameter of 1.85. For the lognormal distribution, the coefficient of variation is .1 for the lightest tail and 1.75 for the heaviest tailed. For distributions with tail heaviness near the extremes or outside of this (vaguely defined) range, the accuracy of the method degenerates at the sample sizes studied.

Although our Monte Carlo results are given only for the Weibull, lognormal and gamma, other results not given show that the accuracy holds up for mixtures of these families. There seems to be two essential requisites for accuracy. First, that the tail heaviness be in a certain range. Second, that the tail decreases smoothly in the extrapolation region.

All bets are off if nature is a trickster. For instance, there are some conjectures that ozone levels in the lower atmosphere have an upper bound imposed by the nature of the chemical processes that scavenge ozone. If so, extrapolations to the neighborhood of this bound based on the quadratic tail model will given erroneous results.

3.2 *Tail heaviness.* We have used the terms heavy-tailed and light-tailed above without precise definitions. In our work we found it convenient to define a measure of tail heaviness so that various distributions could be categorized and compared.

Furthermore, a single number characterizing tail heaviness did not seem either revealing or useful. Instead, we wanted a measure $H_X(p)$ of heaviness of the distribution of the variable X at the quantile x_p.

Two properties were required. First, that our reference distribution, the exponential, have zero tail heaviness. Second, that $H_X(p)$ be invariant under changes of location and scale, i.e. that

$$H_{c+dX}(p) = H_X(p).$$

A convenient measure satisfying these two conditions is given by

DEFINITION 3.1. The tail heaviness of a distribution is

$$H_X(p) = -p \left[\frac{d^2x_p}{dp^2} \Big/ \frac{dx_p}{dp} \right] - 1$$

or, equivalently

$$H_X(p) = \frac{pG''(x_p)}{(G'(x_p))^2} - 1, \quad G(x) = 1 - F(x).$$

If X is exponential, then $H_X(p) = 0$. In the following a distribution with $H_X(p) > 0$ will be called heavy-tailed at x_p and light-tailed if $H_X(p) < 0$.

Suppose X has an exponential distribution. Then $X^{1/\lambda}$ has a Weibull distribution with shape parameter λ. A short computation shows that

$$H_{X^{1/\lambda}}(p) = \frac{1-\lambda}{\lambda \log(1/p)} .$$

Therefore a Weibull is heavy-tailed if $0 < \lambda < 1$ and light-tailed if $\lambda > 1$. Notice that if $\lambda \neq 1$, $H_{X^{1/\lambda}}(p)$ converges to zero very slowly as $p \rightarrow 0$.

Now take X to be lognormal, $X = \exp(\sigma Z)$, with shape parameter $\sigma > 0$, where Z is a unit normal with density $\phi(x)$ and quantiles z_p. By a routine computation

$$H_X(p) = \frac{p(z_p+\sigma)}{\phi(z_p)} - 1 .$$

Using straightforward asymptotics, we get

$$\lim_{p\rightarrow 0} H_X(p)(\log \frac{1}{p})^{1/2} = \sigma/\sqrt{2} .$$

So, for p small enough, the lognormal is always heavy-tailed, but the rate of convergence of $H_X(p)$ to zero is even slower than for a Weibull.

Finally, we note that for the quadratic tail model, $H_X(p)$ has the simple expression

$$H_X(p) = \frac{b}{a+b \log \frac{1}{p}}, \quad p \leq p_0 .$$

if a and b are both positive, the tail heaviness is positive, and decreases to zero as $p \rightarrow 0$ at the same rate as for heavy-tailed Weibulls. If $a > 0$, and $b < 0$, then the tail heaviness is negative for large p and decreases to $-\infty$ at $p = \exp(a/b)$. This suggests that the quadratic tail model may not provide good approximations for extreme quantiles of light-tailed distributions.

In the simulation described, we used the Weibull, generalized gamma, and lognormal distributions with $H_x(.1)$ ranging from -.2 to .4 in steps of .1. The generalized gamma distributions used were of the form $X^{1/\lambda}$ with X a sum of five independent unit exponentials.

To give an intuitive idea of the tail lengths, Table 1 below lists the values of

$$R = \frac{x_{.001} - x_{.5}}{x_{.1} - x_{.5}}$$

for the distributions used. Note that $R = 2.4$ for a normal, and for a Cauchy, $R = 103$.

Table 1
Tail Length Measure R

$H_x(.1)$

Distribution	-.2	-.1	0	.1	.2	.3	.4
Weibull	2.7	3.2	3.9	4.7	5.8	7.2	9.1
Gen. Gamma	2.7	3.2	3.9	4.8	6.0	7.6	9.7
Lognormal	2.7	3.3	4.0	5.0	6.3	8.1	10.5

In more standard terms, for the lognormal, the corresponding coefficients of variation were .12, .34, .50 ,.72, .99, 1.31, .1.72. For the Weibull, the shape parameters were 1.85, 1.30, 1.00, .81, .68, .59, .52. For the generalized gamma, the power parameter λ had the values 1.47, .88, .63, .49, .40, .34, .29.

3.3 *Range of sample sizes, quantiles estimated, and length of tails fitted.* The minimum sample sizes for which the quadratic tail method is reasonably accurate is determined by a trade-off between bias and variance. For a given sample $X_1, \ldots, X_n$ we use the order statistics $X_{(1)}, \ldots, X_{(m)}$ to fit the tail of the distribution. If m is too small, the estimates $\hat{a}, \hat{b}$ will be noisy. If we try to increase m by making it a larger fraction of n, then we risk fitting too much of the bulk of the distribution and introducing substantial bias into the tail estimates. However, our simulations show that the accuracy holds up surprisingly well even for n as small as 50, with $m = 25$. The simulation results are given for $n = 50$ and $n = 200$. Our results for other sample sizes show that for $50 < n < 200$, a good approximation to the accuracy can be gotten by log-linear interpolation between the $n = 50$ and the $n = 200$ results. For $n > 200$, the accuracy, of course, increases.

The appropriate value of m/n varies with n. For n small, m/n must be taken larger to reduce the variability of $\hat{a}, \hat{b}$. For example, at $n = 50$, we use $m = 25$, $m/n = .5$. At larger values of n, m/n can be smaller, thus reducing the bias, while keeping m large enough to keep the variance in control. At $n = 200$, we found that $m = 55$, $m/n \doteq 0.2$ was a satisfactory compromise, and at $n = 800$, $m = 80$, $m/n = .1$ gave good results.

The range of quantiles that can be estimated with reasonable accuracy depends strongly on the sample size n. Roughly speaking, the range of quantiles has to be "within reach of the data". For instance, $X_{(1)}$ is a median unbiased estimator of the quantile x_p, $p \doteq (\log 2)/n$. Thus, quantiles x_p, with $np \lesssim .7$, can be estimated using more or less standard nonparametrics.

We have found that the range of quantiles that can be accurately estimated by the tail quadratic method is also governed by the value of np. Originally, we had hoped to be able to get as far as $np = .1$. To our surprise, for the distributions studied, accuracy is only mildly eroded as far up as $np = .01$.

We describe the results for the np values .01, .1, 1.0. In addition, some runs were made at larger np values to compare the quadratic tail confidence intervals with nonparametric confidence intervals.

4. Simulation Results.

4.1 *General set-up.* Sections 4.3 - 4.5 below give summaries of the bias of the quadratic tail estimator, the actual coverage at the computed confidence bounds, and measures of the size of the computed bounds. Section 4.6 compares the size of the quadratic bounds with nonparametric bounds.

For given values of n and np, $\bar{t}_{.9}$ and $\underline{t}_{.9}$ were derived using 10,000 repetitions in the procedure described at the end of Section 2. All other measures for given n, np tail heaviness and distribution were computed using 10,000 repetitions. These runs were done on the Boeing CRAY I.

4.2 *General format.* In this section we give a number of figures in similar format to graphically describe our simulation results. To avoid repetitions labeling, we note that in all figures, the horizontal axis is the tail heaviness $H_x(.1)$ of the distributions. All figures contain 3 curves giving the results for the Weibull, generalized gamma and lognormal distributions. The distinction is made by

(Solid lines) ______ Weibull

(Dashed lines) - - - - generalized gamma

(Dotted lines) · · · · · · lognormal .

4.3 *Bias.* The issue of bias is critical in estimating tail quantiles. In the simulation, 10,000 repetitions were done for each value of n, np, tail heaviness and distribution, resulting in 10,000 estimates $\hat{x}_{p,1}, \ldots, \hat{x}_{p,10000}$ of x_p. The percent bias was defined as

$$100 \bullet \underset{k}{\text{Avg}} \left(\frac{\hat{x}_{pk} - x_p}{x_p} \right) .$$

The results are given in Figure 1.

[Figure 1 about here.]

4.4 *Coverage.* We describe the accuracy of the confidence bounds only for the 90% bounds. If the bounds were perfectly accurate, then 90% of the time

$$x_p \leq \hat{x}_p + \bar{t}_{.9}\sigma(\hat{a},\hat{b}) \quad \text{(UCB)}$$

and also 90% of the times

$$x_p \geq \hat{x}_p + \underline{t}_{.9}\sigma(\hat{a},\hat{b}) \quad \text{(LCB).}$$

The actual coverage percentages (in the 10,000 repetitions) achieved by these bounds is graphed in Figure 2 for $n = 50$ and in Figure 3 for $n = 200$.

[Figures 2 and 3 about here.]

4.5 *Size of the bounds.* In Section 4.4 the results show that the coverage stays fairly close to the targeted 90% going above 95% only for the short-tailed distributions and below 80% only for the heavy-tailed lognormals. An important issue is how large these bounds are in order to achieve the given coverages. As a measure of this for the upper bounds, for any single run we defined the percent excess as

$$100 \bullet \frac{\text{UCB} - x_p}{x_p}$$

where UCB is the estimated 90% upper confidence bound. The overall percent excess was taken to be the median of the above numbers, over the 10,000 runs. Similarly, for

Figure 1

Percent Bias

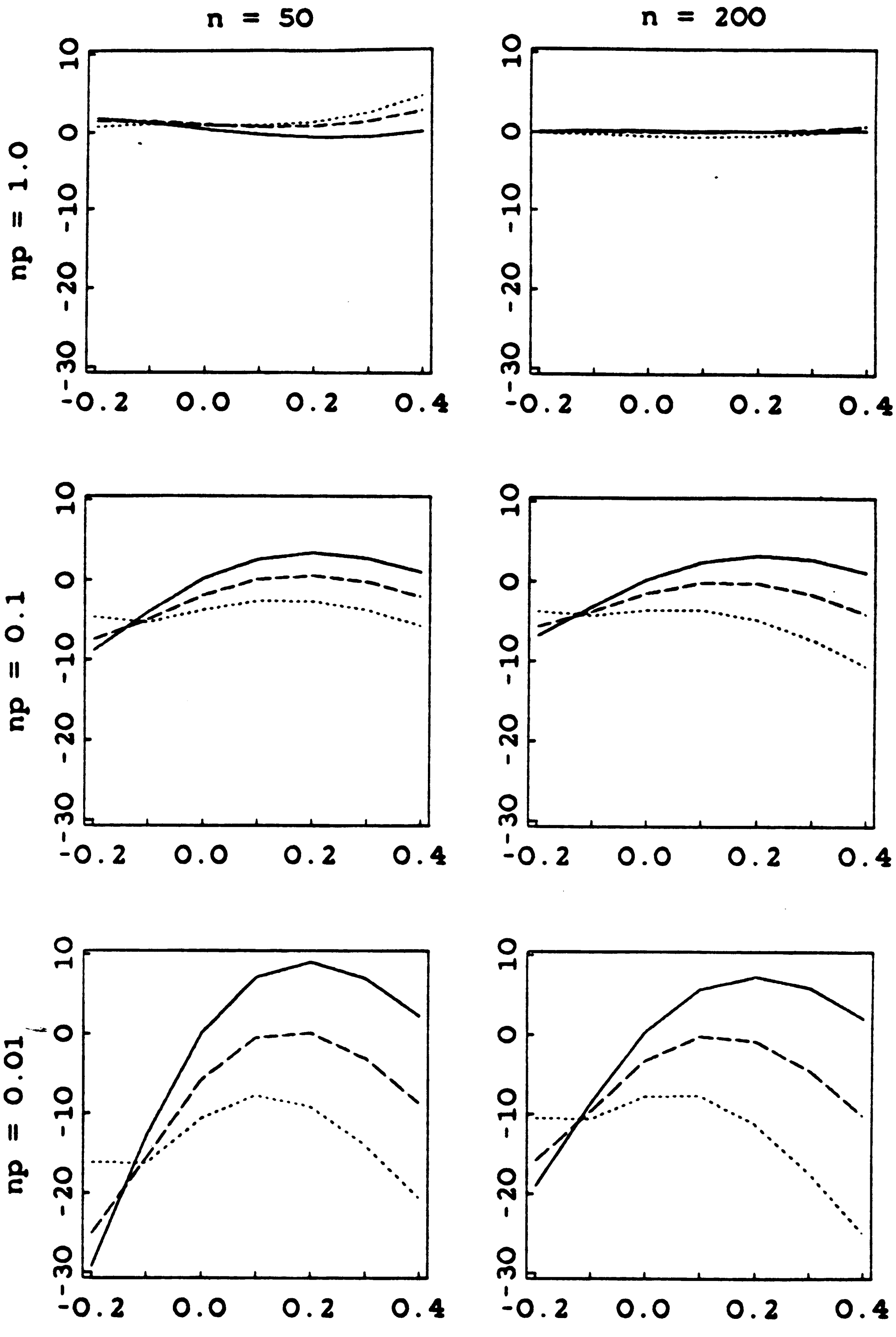


Figure 2

Percent Coverage (n = 50)

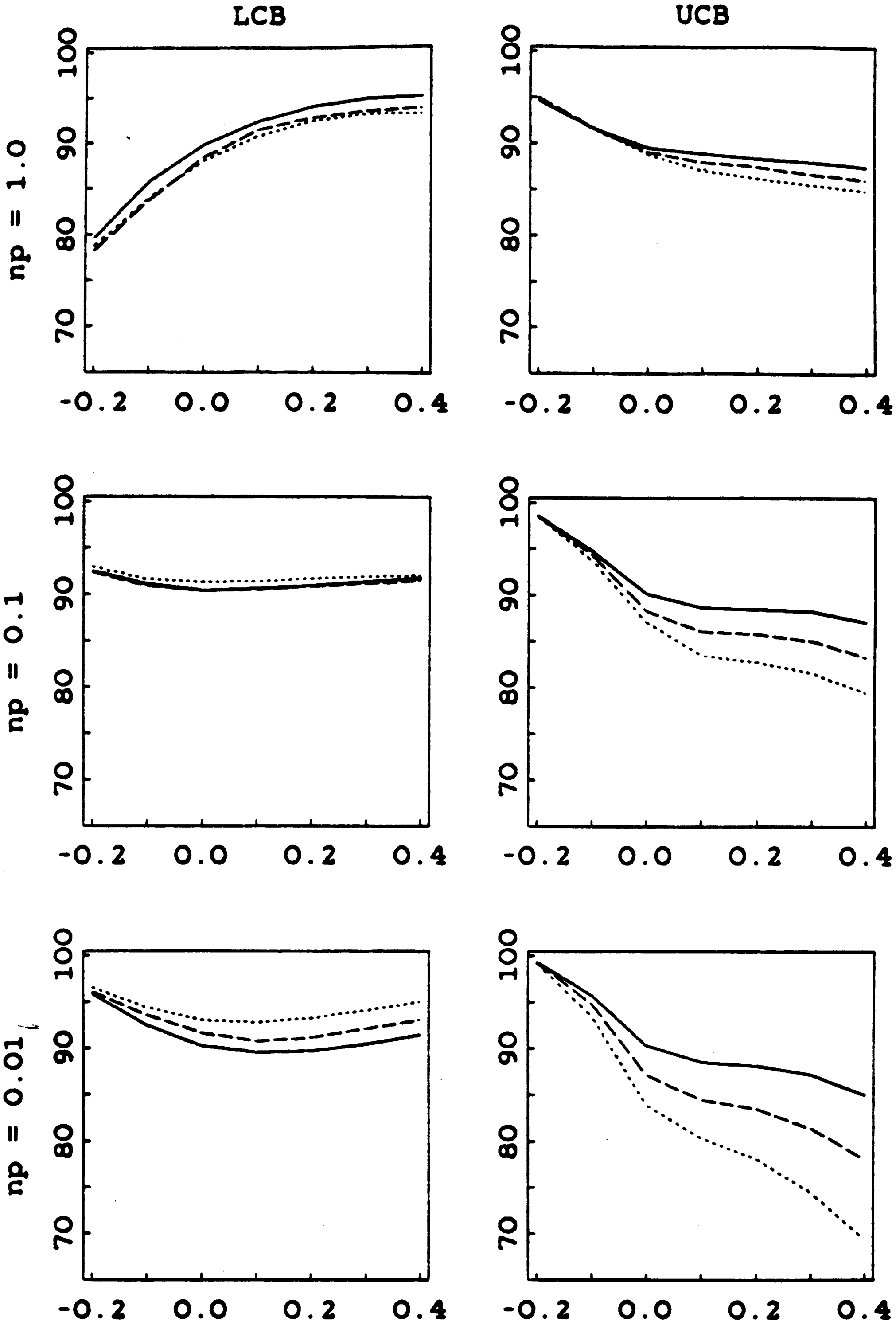
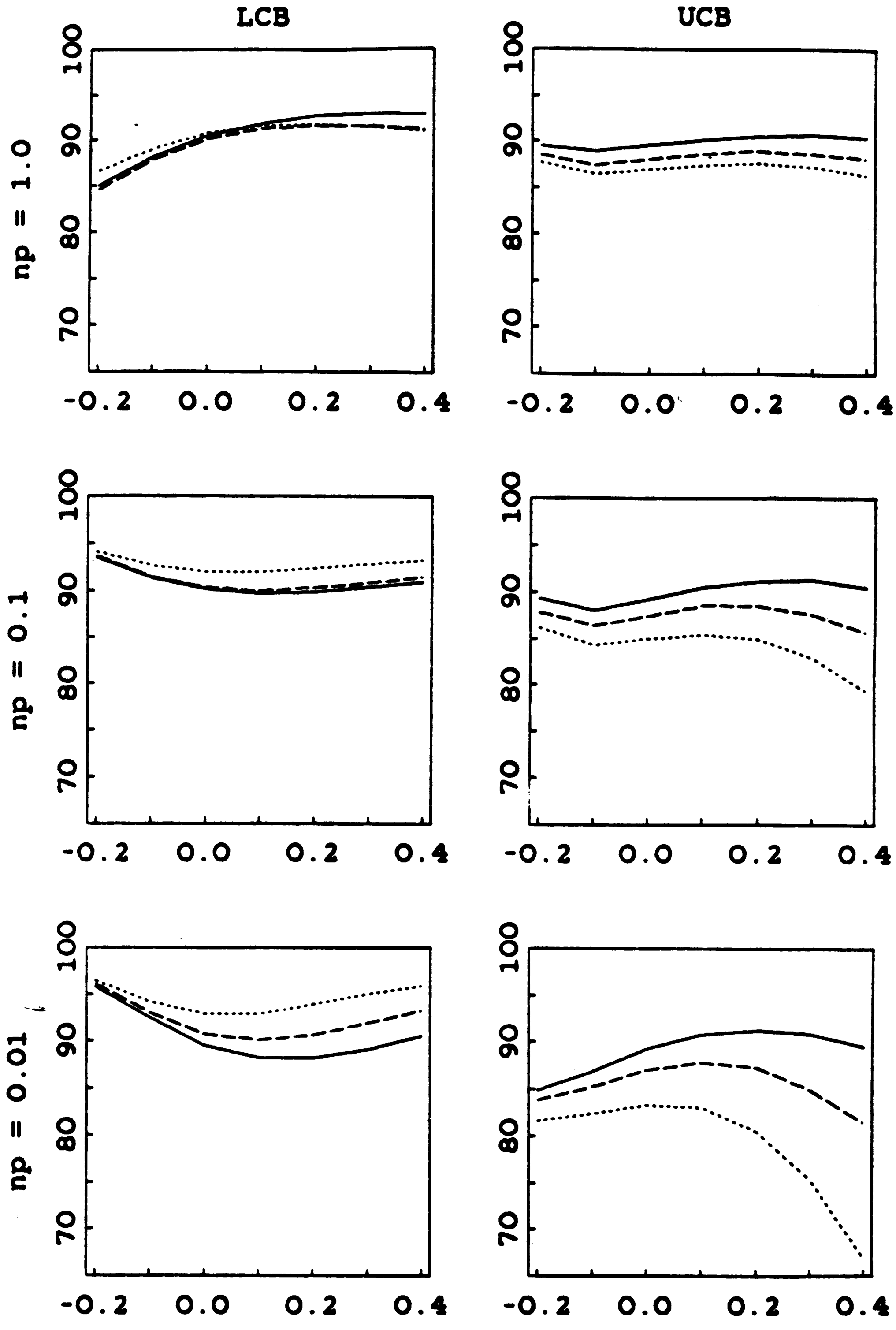


Figure 3

Percent Coverage (n = 200)



the lower bounds, the percent excess was defined as

$$100 \bullet \frac{x_p - LCB}{x_p}$$

and the median over 10,000 runs computed.

The results are graphed in Figure 4 for $n = 50$ and in Figure 5 for $n = 200$.

[Figures 4 and 5 about here.]

4.6 *Comparison with nonparametric bounds.* Looking at the percent excess graphed in Section 4.5, one is struck by the fact that often the upper confidence bound is over 50% larger than the quantile being estimated. Since anyone can get good coverage by using large enough bounds, this raises the question of whether the large size of these bounds is in the nature of the problem, or is reflective of inefficiency in the quadratic tail bounds.

Following a suggestion due to Peter Bickel, we explored this issue as follows: for each of $n = 50$, $n = 200$ select p such that for $X_{(1)}$ the largest order statistic,

$$P(x_p \leq X_{(1)}) = .9 .$$

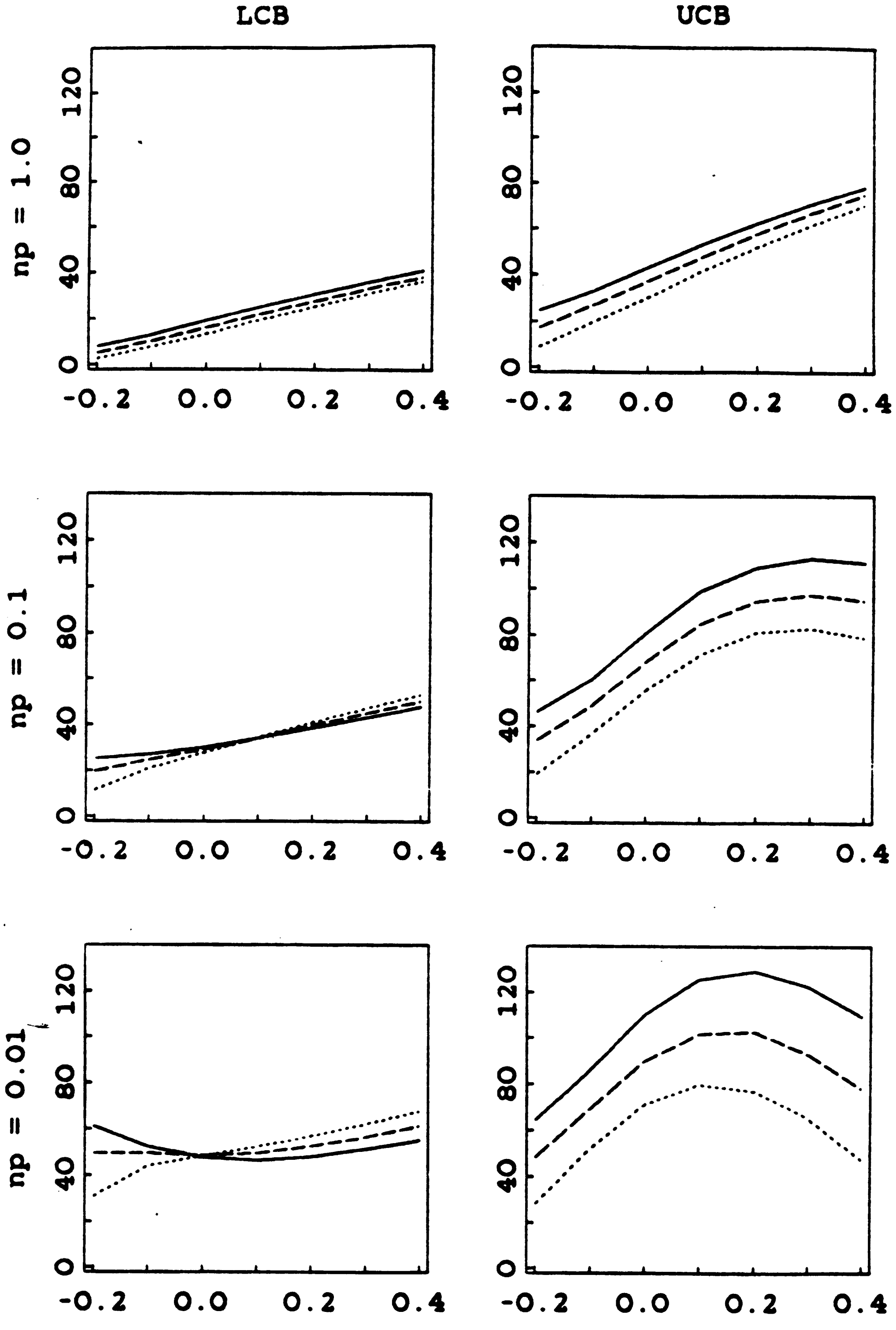
Thus, $X_{(1)}$ is a nonparametric 90% upper confidence bound for x_p. Over 10,000 repetitions, compute the median of

$$100 \bullet \frac{X_{(1)} - x_p}{x_p} .$$

This gives the same measure of percent excess as that computed in Section 4.5. Define

Figure 4

Percent Excess (n = 50)

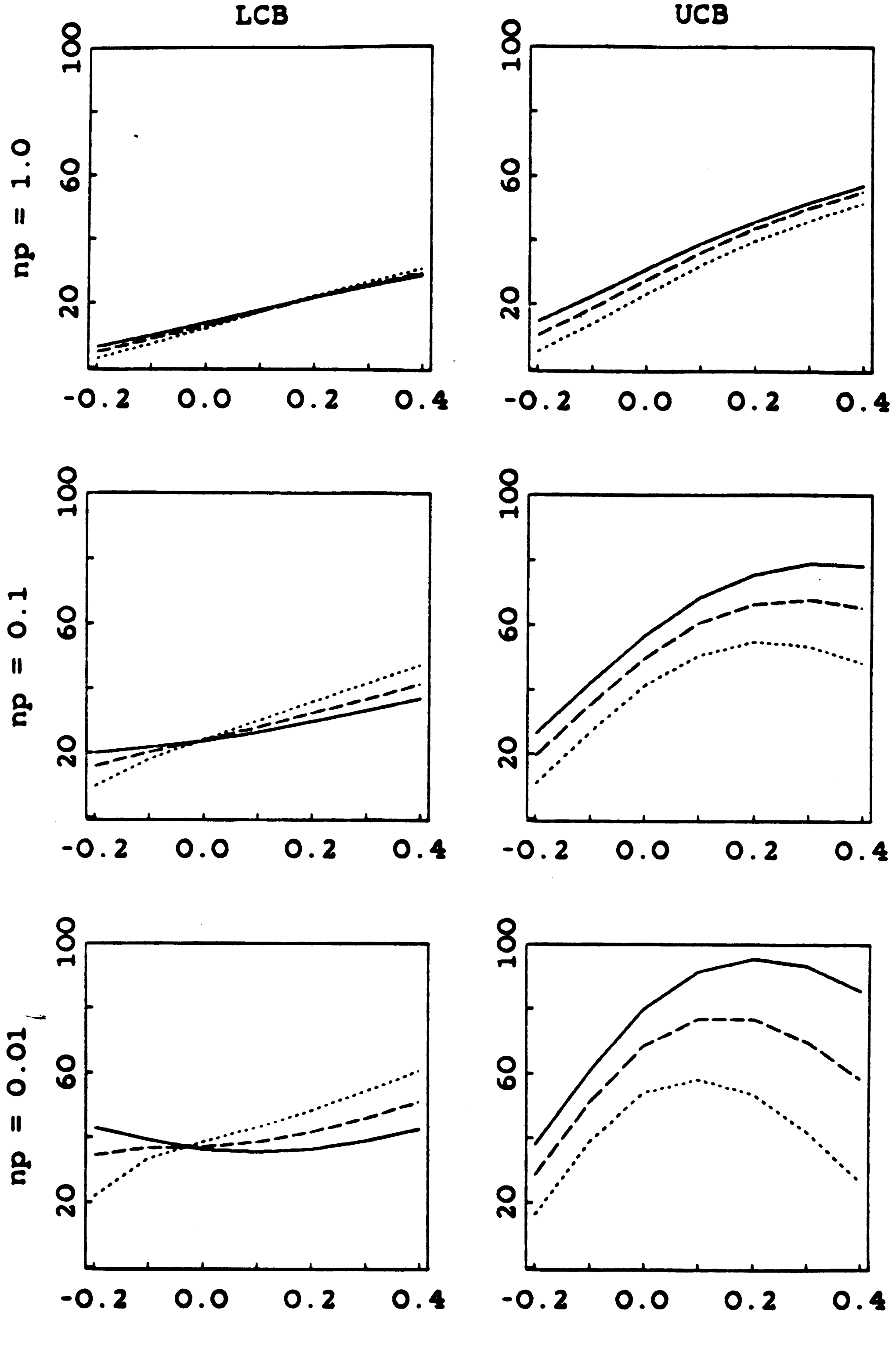


LCB

UCB

Figure 5

Percent Excess (n = 200)



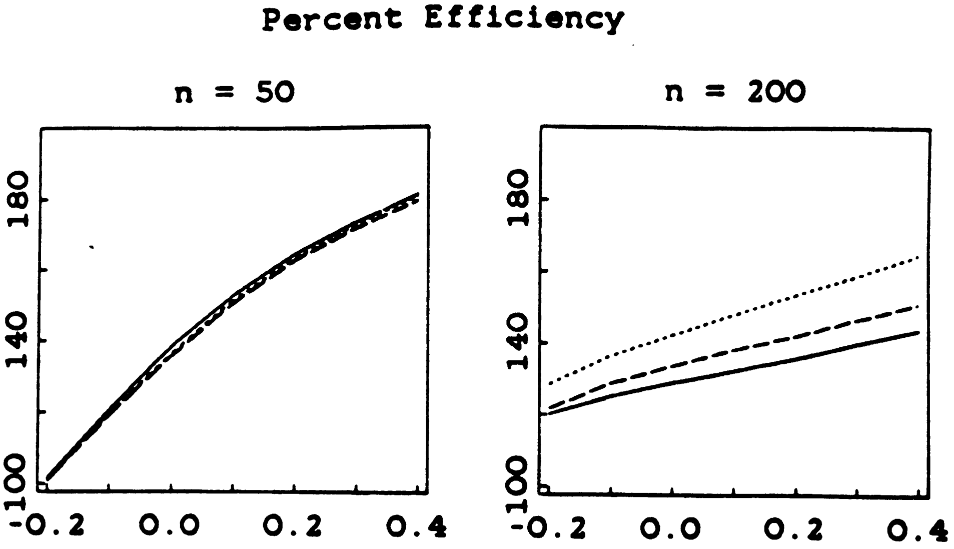
the efficiency of the quadratic tail bound as

$$100 \bullet \frac{\text{percent excess (nonparametric)}}{\text{percent excess (quadratic tail)}} .$$

This efficiency measure is graphed in Figure 6. The efficiency is always at least 100% and is frequently above 150%, doing especially well for the long-tailed distributions. The semi-parametric character of the quadratic tail approach thus provides uniformly smaller bounds over the range of distributions tested than the nonparametric bounds.

[Figure 6 about here.]

Figure 6



5. Remarks and Conclusions. The preceding simulation helps to delineate the range of accuracy of the quadratic tail model. At tail heaviness of .4 and -.2, the coverage probabilities are going awry for the more extreme quantiles. For sample size 50, the results are not as accurate as for the larger sample sizes. However, within these boundaries the quadratic tail bounds have coverage probabilities fairly close to the 90% target.

The price one pays for the broad spectrum of applicability is in the large size of the intervals. But to attempt to get shorter intervals by fitting a global parametric model is a venture based on the magical hope that somehow, the tail distribution of a phenomenon, concerning which one has little or no data, will suitably coincide with the tail distribution of the selected model. Without some spectacular luck (or insight), the results can be disastrously misleading. The "safer" confidence bounds produced by the quadratic tail model, when applicable, make it much more preferable in practice.

The quadratic tail model can be also used to compute confidence intervals for tail parameters other than quantiles. For example, suppose that the parameter of interest is $\tau = E[\max(X_1, \ldots, X_N)]$, the expected maximum of N samples, where N is usually larger than the number n of observations. Now

$$\tau = \int_0^\infty x d(F^N(x)) .$$

so substituting $F(x) = 1 - p$, then $x = x_p$ gives

$$\tau = -\int_0^1 x_p d(1-p)^N .$$

If the quadratic tail model holds in the range $0 < p \leq p_0$, then

$$\tau = La + Mb + R ,$$

where $0 \leq R \leq x_{p_0}(1-p_0)^N$, and

$$L = -\int_0^{p_0} \log(p_0/p) d(1-p)^N \doteq -\int_0^1 \log(p_0/p) d(1-p)^N$$

$$M = \frac{1}{2}\int_0^{p_0}[\log^2 p_0 - \log^2 p]d(1-p)^N \doteq \frac{1}{2}\int_0^1 [\log^2 p_0 - \log^2 p]d(1-p)^N .$$

Therefore,

$$L \doteq \log p_0 + \sum_1^N \frac{1}{k}$$

$$M \doteq -\frac{1}{2}\log^2 p_0 + \sum_1^N \frac{1}{k^2} + (\sum_1^N \frac{1}{k})^2 ,$$

and neglecting the extremely small errors in the approximations, we again have the problem of constructing confidence bounds for a parameter of the form $La + Mb$, with L and M known.

The quadratic tail model is not a "unique extension" of the exponential tail model. Other two-parameter extensions have been investigated. An initially promising approach was to take a power transformation of the data, using that power which made the tail of the transformed data most nearly exponential. Following the transformation an exponential tail model was used. This method was discarded because its accuracy was inferior to that of the quadratic tail procedure.

There are two interesting questions which we leave open. The first is: how well will exponential or quadratic tail procedures work with autocorrelated data? The one piece of evidence we have is that the exponential tail procedure gave quite accurate estimates of the expected maximum for data generated by an autoregressive scheme with a superimposed trend (see Breiman et al., 1979).

The second question is more tantilizing. Workers often study many data sets drawn from similar sources and have strong opinions regarding their tail heaviness. For instance, one EPA colleague, who has been extensively involved in automotive emissions testing, believes that measured emissions from a sample of automobiles of the same model will have tail heaviness about the same as that of a lognormal with coefficient variation equal to 1.0. This lead us to the hope that by somehow combining the information from many data sets having similar tail heaviness, one could produce more accurate coverage probabilities for the quantiles corresponding to individual data sets. However, our efforts along this line has not been successful, and we leave it as a promising area for future research.

REFERENCES

BOOS, D. D. (1984), "Using Extreme Value Theory to Estimate Large Percentiles, " *Technometrics,* 26, 33-39.

BREIMAN, L., GINS, J. D., and STONE, C. J. (1978), "Statistical Analysis and Interpretation of Peak Air Pollution Measurements," TSC-PD-A190-10, Technology Service Corporation, Santa Monica, California.

BREIMAN, L., STONE, C. J., and GINS, J. D. (1979), "New Methods for Estimating Tail Probabilities and Extreme Value Distribution," TSC-PD-A226-1, Technology Service Corporation, Santa Monica, California

BREIMAN, L., STONE, C. J., and GINS, J. D. (1981), "Further Development of New Methods for Estimating Tail Probabilities and Extreme Value Distributions," TSC-PD-A243-1, Technology Service Corporation, Santa Monica, California.

CRAGER, M. R. (1982), "Exponential Tail Quantile Estimators for Air Quality Data," Technical Report Nos. 4 (Part I) and 5 (Part II), Bay Area Air Quality Management District, San Francisco, California.

CSÖRGŐ, S., DEHEUVELS, P. and MASON, D. (1985), "Kernel Estimates of the Tail Index of a Distribution," *Annals of Statistics,* to appear.

DAVIS, R. and RESNICK, S. (1984), "Tail Estimates Motivated by Extreme Value Theory," *Annals of Statistics,* 12, 1467-1487.

DuMOUCHEL, W. H. (1983), "Estimating the Stable Index α in Order to Measure Tail Thickness: a Critique," *Annals of Statistics,* 11, 1019-1031.

DUMOUCHEL, W. H. and OLSHEN, R. A. (1975), "On the Distribution of Claims Costs," *Credibility,* New York: Academic Press, 23-50.

HALL, P. and WELSH, A. H. (1985), "Adaptive Estimates of Parameters of Regular Variation," ***Annals of Statistics,*** 13, 331-341.

HILL, B. M. (1975), "A Simple General Approach to Inference about the Tail of a Distribution," ***Annals of Statistics,*** 3, 1163-1174.

MARITZ, J. S. and MUNRO, A. H. (1967), "On the Use of the Generalized Extreme-value Distribution in Estimating Extreme Percentiles," ***Biometrics,*** 23, 79-103.

PICKANDS, J. III (1975), "Statistical Inference Using Extreme Order Statistics," ***Annals of Statistics,*** 3, 119-131.

WEISSMAN, I. (1978), "Estimation of Parameters and Large Quantiles Based on the k Largest Observations," ***Journal of the American Statistical Association,*** 73, 812-815.

APPENDIX

In this section (2.9) will be verified and formulas for the first two terms occurring on the right side of (2.10) will be obtained. To this end, recall that $Z_1, \dots, Z_n$ are independent random variables each having an exponential distribution with mean one. The following facts are easily checked: $\mathrm{Var}(Z_1) = 1$; $\mathrm{Var}(Z_1^2) = 20$; $\mathrm{Var}(Z_1 Z_2) = 3$; $\mathrm{Cov}(Z_1, Z_1^2) = 4$; $\mathrm{Cov}(Z_1, Z_1 Z_2) = 1$; $\mathrm{Cov}(Z_1^2, Z_1 Z_2) = 4$; and $\mathrm{Cov}(Z_1 Z_2, Z_1 Z_3) = 1$. It can be assumed that

$$Z_{(k)} = \sum_k^n \frac{Z_j}{j}, \quad 1 \leq k \leq n .$$

Also define δ_{kl} to be 1 if $k = l$ and 0 otherwise; and define ψ_{kl} to be 1 if $k > l$ and 0 otherwise. It is easily seen that

$$\mathrm{Cov}(Z_k, Z_l Z_{(l+1)}) = \delta_{kl} u_l + \frac{\Psi_{kl}}{k} - \frac{\delta_{kl}}{k} ;$$

$$\mathrm{Cov}(Z_k^2, Z_l Z_{(l+1)}) = 4(\delta_{kl} u_l + \frac{\Psi_{kl}}{k} - \frac{\delta_{kl}}{k}) ;$$

$$\mathrm{Var}(Z_k Z_{(k+1)}) = u_k^2 - 2\frac{u_k}{k} + 2u_k^{(2)} - \frac{1}{k^2} ;$$

and

$$\mathrm{Cov}(Z_k Z_{(k+1)}, Z_l Z_{(l+1)}) = u_k^{(2)} + \frac{u_k}{k} - \frac{2}{k^2}, \quad k > l .$$

In verifying (2.9) it can be assumed that

$$\hat{r} = \Sigma\, k w_k [a(Z_{(k)} - Z_{(k+1)}) + \frac{b}{2}(Z_{(k)}^2 - Z_{(k+1)}^2)] = a\,\Sigma\, w_k Z_k + b\,\Sigma\, w_k (\frac{Z_k^2}{2k} + Z_k Z_{(k+1)})$$

(with k ranging from 1 to $m-1$ unless otherwise indicated). Write $\mathrm{Var}(\hat{r}) = B_1 a^2 + 2B_2 ab + B_3 b^2$. Then

$$B_1 = \mathrm{Var}(\Sigma\, w_k Z_k) = \Sigma\, w_k^2$$

and

$$B_2 = \mathrm{Cov}\Big(\Sigma\, w_k Z_k,\ \Sigma\, w_k\Big(\frac{Z_k^2}{2k} + Z_{(k)}Z_{(k+1)}\Big)\Big)$$

$$= 2\Sigma\,\frac{w_k^2}{k} + \Sigma\,\Sigma\, w_k w_l\,\Big(\delta_{kl}u_l + \frac{\psi_{kl}}{k} - \frac{\delta_{kl}}{k}\Big)$$

$$= \Sigma w_k(\bar{w}_k + u_k w_k)\,.$$

Also,

$$B_3 = \mathrm{Var}\Big(\frac{1}{2}\,\Sigma\frac{w_k}{k}\,Z_k^2 + \Sigma w_k Z_k Z_{(k+1)}\Big) = B_4 + B_5 + B_6\,.$$

Here

$$B_4 = \frac{1}{4}\mathrm{Var}\Big(\Sigma\frac{w_k}{k}\,Z_k^2\Big) = 5\,\Sigma\frac{w_k^2}{k^2}\,.$$

Next,

$$B_5 = \mathrm{Cov}\Big(\Sigma\frac{w_k}{k}\,Z_k^2,\ \Sigma\, w_k Z_k Z_{(k+1)}\Big)$$

$$= 4\,\Sigma\frac{w_k}{k}\,\Sigma\, w_l\,\Big(\delta_{kl}u_l + \frac{\psi_{kl}}{k} - \frac{\delta_{kl}}{k}\Big)$$

$$= 4\,\Sigma\,\frac{u_k w_k^2}{k} + 4\,\Sigma\,\frac{w_k\bar{w}_k}{k} - 8\,\Sigma\,\frac{w_k^2}{k^2}\,.$$

Further,

$$B_6 = \mathrm{Var}(\Sigma\, w_k Z_k Z_{(k+1)})$$

$$= \Sigma\, w_k^2 (u_k^2 - 2\frac{u_k}{k} + 2\,u_k^{(2)} - \frac{1}{k^2}) + 2\,\Sigma\, w_k (u_k^{(2)} + \frac{u_k}{k} - \frac{2}{k^2})(k\bar{w}_k - w_k)$$

$$= \Sigma\, u_k^2 w_k^2 - 4\,\Sigma\, \frac{u_k w_k^2}{k} + 3\,\Sigma \frac{w_k^2}{k^2} + 2\,\Sigma\, k u_k^{(2)} w_k \bar{w}_k$$

$$- 4\,\Sigma\, \frac{w_k \bar{w}_k}{k} + 2\,\Sigma\, u_k w_k \bar{w}_k\,.$$

Consequently,

$$B_3 = \Sigma\, u_k^2 w_k^2 + 2\,\Sigma\, k u_k^{(2)} w_k \bar{w}_k + 2\,\Sigma\, u_k w_k \bar{w}_k\,.$$

Observe that

$$\Sigma\, \bar{w}_k^2 = \Sigma\, (u_k^{(2)} - u_{k+1}^{(2)}) (\Sigma_1^k\, w_l)^2 = 2\,\Sigma k u_k^{(2)} w_k \bar{w}_k - \Sigma\, u_k^{(2)} w_k^2 - u_m^{(2)} ((m-1)\bar{w}_{m-1})^2$$

and hence that

$$B_3 = \Sigma\, (u_k w_k + \bar{w}_k)^2 + \Sigma\, u_k^{(2)} w_k^2 + u_k^{(2)} ((m-1)\bar{w}_{m-1})^2\,.$$

The last formula for B_3 and the previous formulas for B_1 and B_2 together show that (2.9) is valid.

We will now determine formulas for the terms $\mathrm{Var}(X_{(m)})$ and $\mathrm{Cov}(X_{(m)}, \hat{r})$ appearing in the right side of (2.10). First,

$$\mathrm{Var}(X_{(m)}) = \mathrm{Var}(aZ_{(m)} + \frac{b}{2}\, Z_{(m)}^2)$$

and hence

$$\mathrm{Var}(X_{(m)}) = a^2\, \mathrm{Var}(Z_{(m)}) + ab\mathrm{Cov}(Z_{(m)}, Z_{(m)}^2) + \frac{b^2}{4}\, \mathrm{Var}(Z_{(m)}^2)\,. \qquad \text{(A.1)}$$

It will be shown below that

$$\mathrm{Var}(Z_{(m)}) = u_m^{(2)}\,, \qquad \text{(A.2)}$$

$$\mathrm{Cov}\,(Z_{(m)}, Z^2_{(m)}) = 2(u^{(3)}_m + u^{(2)}_m u_m)\,, \qquad (A.3)$$

and

$$\mathrm{Var}\,(Z^2_{(m)}) = 6u^{(4)}_m + 8u^{(3)}_m u_m + 2\,(u^{(2)}_m)^2 + 4u^{(2)}_m u^2_m\,, \qquad (A.4)$$

where

$$u^{(3)}_m = \sum_m^n \frac{1}{j^3} \quad \text{and} \quad u^{(4)}_m = \sum_m^n \frac{1}{j^4}\,.$$

Equations (A.1) - (A.4) together yield the desired formula for $\mathrm{Var}\,(X_{(m)})$. Secondly,

$$\begin{aligned}\mathrm{Cov}\,(X_{(m)}, \hat{r}) &= \mathrm{Cov}\,(aZ_{(m)} + \frac{b}{2}\,Z^2_{(m)}, b\,\Sigma\, kw_k Z_{(k+1)}\,(Z_{(k)} - Z_{(k+1)}) \\ &= \mathrm{Cov}\,(aZ_{(m)} + \frac{b}{2}\,Z^2_{(m)}, b\,(\Sigma\, w_k)Z_{(m)})\end{aligned}$$

and hence

$$\mathrm{Cov}\,(Z_{(m)}, \hat{r}) = (m-1)\,\overline{w}_{m-1}(ab\mathrm{Var}\,(Z_{(m)}) + \frac{b^2}{2}\,\mathrm{Cov}\,(Z_{(m)}, Z^2_{(m)})\,. \qquad (A.5)$$

Equations (A.2), (A.3), and (A.5) together determine the desired formula for $\mathrm{Cov}\,(Z_{(m)}, \hat{r})$.

It remains to verify (A.2) - (A.4). To this end, let i, j, k, l range from m to n. Then

$$\mathrm{Var}\,(Z_{(m)}) = \mathrm{Var}\,(\Sigma\,\frac{Z_j}{j}) = \Sigma\,\frac{1}{j^2} = u^{(2)}_m\,,$$

so (A.2) holds. Observe next that

$$\begin{aligned}
\mathrm{Cov}(Z_{(m)}, Z_{(m)}^2) &= \mathrm{Cov}(\Sigma \frac{Z_j}{j}, (\frac{Z_j}{j})^2) \\
&= \Sigma\,\Sigma\,\Sigma \frac{1}{jkl} \mathrm{Cov}(Z_j, Z_k Z_l) \\
&= \Sigma \frac{1}{j^3} \mathrm{Cov}(Z_j, Z_j^2) + 2\,\Sigma \frac{1}{j^2} \sum_{k \neq j} \frac{1}{k} \mathrm{Cov}(Z_j, Z_j Z_k) \\
&= 4u_m^{(3)} + 2\,\Sigma \frac{1}{j^2} (\Sigma \frac{1}{k} - \frac{1}{j}) \\
&= 2\,(u_m^{(3)} + u_m^{(2)} u_m),
\end{aligned}$$

so (A.3) holds. Finally,

$$\mathrm{Var}(Z_{(m)}) = \mathrm{Var}((\Sigma \frac{Z_j}{j})^2) = \Sigma\,\Sigma\,\Sigma\,\Sigma \frac{1}{ijkl} \mathrm{Cov}(Z_i Z_j, Z_k Z_l).$$

The total contribution of all terms for which $i = j = k = l$ is

$$\mathrm{Var}(Z_1^2)\,\Sigma \frac{1}{j^4} = 20 u_m^{(4)}.$$

The total contribution of all terms for which exactly three of the four quantities i, j, i, l coincide is

$$4\mathrm{Cov}(Z_1^2, Z_1 Z_2) \Sigma \frac{1}{j^3} \sum_{k \neq j} \frac{1}{k} = 16\,(u_m^{(3)} u_m - u_m^{(4)}).$$

The total contribution of all terms for which i and j are distinct and exactly one of the pair k, l equals either i or j is

$$\begin{aligned}
4\mathrm{Cov}(Z_1 Z_2, Z_1 Z_3) \underset{\substack{j,k,l \\ \text{distinct}}}{\Sigma\,\Sigma\,\Sigma} \frac{1}{j^2 k l} &= 4\,\Sigma \frac{1}{j^2} \sum_{k \neq j} \frac{1}{k} (u_m - \frac{1}{j} - \frac{1}{k}) \\
&= 8u_m^{(4)} - 8u_m^{(3)} u_m - 4(u_m^{(2)})^2 + 4u_m^{(2)} u_m^2.
\end{aligned}$$

The total contribution of all terms for which i and j are distinct and (k,l) is either (i,j) or (j,i) is

$$2\mathrm{Var}(Z_1Z_2)\,\Sigma\,\frac{1}{j^2}\sum_{k\neq j}\frac{1}{k^2} \;=\; 6\,(\,(u_m^{(2)})^2 - u_m^{(4)})\,.$$

Equation (A.4) follows by summing these four totals.