To obtain the conditional expected value and conditional variance of

Var $\hat{m}(x_i)$ we utilize the following result:

If for any expression Z, E[Z/U]=A(U)+O(B) and Var[Z/U]=O/C, then $Z=A(U)+O_p(B+C^{1/2})$.

Using the above result we obtain

$$\operatorname{Var} \hat{m}(x_i) = (n-1)^{-1} \left\{ \begin{array}{l} \sigma^2(x_i)+b^2\sigma^2(x_i)k_2 d_s(x_i) d_s^{//}(x_i) \\ +O\left(b^3+(n-1)^{-1/2}b^{-1/2}\right) \end{array} \right\}$$

as $b \to 0$,

$$\operatorname{Var} \hat{m}(x_i) = (n-1)^{-1}\left\{\sigma^2(x_i)+O\left((n-1)^{-1/2}b^{-1/2}\right)\right\}.$$

$$=(n-1)^{-1}\{\sigma^2(x_i)+O(nb)^{-1/2}\}$$

Therefore

$$VarR_i = \sigma^2(x_i)+\frac{\sigma^2(x_i)}{n-1}+O\left(\frac{1}{n^{3/2}b^{1/2}}\right).$$

As $n \to \infty$, and as $(nb)^{1/2} \to \infty$,

$Var(R_i)= \sigma^2(x_i)$ which is the same as the variance of e_i in model (2.2).

REFERENCES.

Chambers, R. L. and Dorfman, A. H. (1994). Robust sample survey inference via bootstrapping and bias correction: The case of the ratio estimator. *Paper presented at the joint meeting of American Statistical Association, Biometrics Society and Statistical Society, Canada.*

Cochran, W. G. (1977). Sampling techniques (3rd ed.), New York: John Wiley.

Dorfman, A. H. (1992). Non-parametric regression for estimating totals infinite populations. *Proceedings of a section on survey research methods. American Statistical Association.* 622-625.

Liu, R. Y. (1988). Bootstrap procedures under some non-IID Models. *Annals of Statistics,* **16,** 1696-1708.

The East African Journal of Statistics ISSN: 1811-7503
Volume1, Number 1, pp. 91-108

ASYMPTOTIC LINEAR ESTIMATION OF THE QUANTILE FUNCTION OF A LOCATION-SCALE FAMILY OF DISTRIBUTIONS BASED ON SELECTED ORDER STATISTICS

PATRICK G. O. WEKE

Department of Mathematics,
University of Nairobi,
P.O. Box 30197, 00100,
Nairobi, Kenya.
Email: *pweke@uonbi.ac.ke*

Abstract. Some general asymptotic methods of estimating the quantile function, $Q(\xi)$, $0<\xi<1$, of location-scale families of distributions based on a few selected order statistics are considered, with applications to some non-regular distributions. Specific results are discussed for the ABLUE of $Q(\xi)$ for the location-scale exponential and double exponential distributions. As a further application of the exponential results, we discuss the asymptotically best optimal spacings for the location-scale logistic distribution.

Key words and phrases: Quantiles; Order statistics; Optimal spacing; Exponential; Logistic.

1. **Introduction.** In the location-scale model it is assumed that the distribution function (df) for the random sample $X_1, X_2, \ldots, X_n$, has the form

$$F(x)=F_0\left(\frac{x-\mu}{\sigma}\right) \tag{1.1}$$

where F_0 is a known distribution function and μ and σ are the unknown location and scale parameters respectively. It is assumed that F is absolutely continuous with respect to Lebesgue measure and has the probability density function (pdf), f, given by

$$f(x)=\frac{1}{\sigma}f_0\left(\frac{x-\mu}{\sigma}\right), \tag{1.2}$$

where $F_0' = f_0$. The quantile function is defined as

$$Q(\xi)=F^{-1}(\xi). \tag{1.3}$$

By the location-scale property we have

$$Q(\xi)=\mu+\sigma Q_0(\xi), \quad (1.4)$$

where $Q_0(\xi)=F_0^{-1}(\xi)$.

A natural estimator of $Q(\xi)$ is the sample quantile function, $\dot{Q}_n(\xi)$, which is the inverse of the empirical distribution function $\dot{F}_n(x)$, defined by

$$\dot{F}_n(x)=\begin{cases}0 & ,x<X_{(1)}\\ (j-1)/n & ,X_{(j-1)}\le x<X_{(j)}\\ 1 & ,x\ge X_{(n)}\end{cases} \quad (1.5)$$

where $X_{(1)},X_{(2)},....,X_{(n)}$, are the order statistics corresponding to the random sample. It is easily seen that

$$\dot{Q}_n(\xi)=X_{(j)} \quad \text{for } (j-1)/n<\xi\le j/n,\ j=1,.....,n, \quad (1.6)$$

which is a piecewise constant function.

Parzen (1979) suggests the following piecewise linear estimator, $\tilde{Q}_n(\xi)$, as preferable to $\dot{Q}_n(\xi)$:

$$\tilde{Q}_n(\xi)=n(j/n-\xi)X_{(j-1)}+n\left(\xi-\frac{j-1}{n}\right)X_{(j)}, \quad (1.7)$$

for $(j-1)/n\le\xi\le j/n$ and $j=1,2,.....,n$, where $X_{(0)}$ is taken as suggested by Parzen (1979).

Dixon (1957, 1960) proposed the simplified linear estimators for the mean and standard deviation of the normal population in terms of the sample quasi-midrange and quasi-range, respectively. Sarhan and Greenberg (1962) give the BLUE of $Q(\xi)$ for the two parameter exponential distribution in complete samples, while Epstein (1960) considered the one parameter case.

Hassanein (1968, 1972) considered the estimation of $Q(\xi)$ for the Gumbel distributin for the large samples and Mann and Fertig (1977) for moderate samples. For small to moderate samples Mann (1970) gives estimators of $Q(0.1)$ and $Q(0.05)$ for the first extreme-value distribution. Kubat and Epstein (1980) considered the estimation of $Q(\xi)$ based on two or three order statistics for the normal and Gumbel distributions. Ali, Umbach, and Hassanein (1981) followed the approach of Kubat and Epstein (1980) for the exponential and double-exponential distributions based on two selected order statistics.

In this paper, we discuss a general theory to obtain ABLUE of $Q(\xi)$ based on k $(\le n)$ selected order statistics from a location-scale family of distributions. This approach, contrary to the approaches taken by Kubat and Epstein (1980) and Ali et al. (1981) enables us to generalize to any arbitrary number of selected order statistics. As an application of this theory, ABLUE of $Q(\xi)$ is obtained

for the location-scale exponential and logistic distributions models and follows closely the work of Balakrishnan and Kannan (2001) and Weke (2001).

2. **ABLUE of quantiles in location-scale families.** We consider estimation of the population quantile $Q(\xi)$ in model (1.1) based on k $(\le n)$ order statistics, $X_{(n_1)}<X_{(n_2)}<.....<X_{(n_k)}$, where $n_i=[np_i]+1$, $i=1,2,....,k$, and $[.]$ is the greatest integer function. The $k-$tuple, $(p_1,p_2,....,p_k)$ whose elements satisfy $0<p_1<p_2<....<p_k<1$ is called a spacing for the sample quantiles. Our goal is to choose the spacing optimally to obtain ABLUE of $Q(\xi)$ based on k selected order statistics.

Given a fixed $k-$tuple, $(p_1,p_2,....,p_k)$, the corresponding sample quantiles $\{\dot{Q}_n(p_i),i=1,2,.....,k\}$, have a $k-$variate asymptotic normal distribution (Mosteller, 1946) with mean vector $(\mu+\sigma Q_0(p_1),...,\mu+\sigma Q_0(p_k))$ and covariance matrix with elements

$$\frac{\sigma^2 p_i(1-p_j)}{nd_0(p_i)d_0p_j} \quad \text{for } 1\le i\le j\le n, \quad (2.1)$$

where $d_0(p_i)=f_0(Q_0(p_i))$ is the density-quantile function at p_i, $i=1,2,...,k$. The ABLUE of the quantile function, $Q(\xi)$, with fixed spacing $(p_1,p_2,....,p_k)$, can be obtained through the generalized least-squares principle as follows:

$$\hat{Q}(\xi)=\frac{1}{\Delta}\{(K_2X+Q_0(\xi)K_1Y)-K_3(Q_0(\xi)X)\}, \quad (2.2)$$

$$\text{where} \quad \Delta=K_1K_2-K_3^2, \quad (2.3)$$

$$X=\sum_{i=1}^{k+1}\left\{\frac{d_0(p_i)-d_0(p_{i-1})}{p_i-p_{i-1}}-\frac{d_0(p_{i+1})-d_0(p_i)}{p_{i+1}-p_i}\right\}d_0(p_i)X_{(n_i)}, \quad (2.4)$$

$$Y=\sum_{i=1}^{k+1}\left\{\frac{d_0(p_i)d_0(p_i)-d_0(p_{i-1})d_0(p_i)}{p_i-p_{i-1}}\right.$$

$$\left.-\frac{Q_0(p_{i+1})d_0(p_{i+1})-Q_0(p_i)d_0(p_i)}{p_{i+1}-p_i}\right\}d_0(p_i)X_{(n_i)}, \quad (2.5)$$

$$K_1=\sum_{i=1}^{k+1}\frac{\{d_0(p_i)-d_0(p_{i-1})\}^2}{p_i-p_{i-1}}, \qquad (2.6a)$$

$$K_2=\sum_{i=1}^{k+1}\frac{\{Q_0(p_i)d_0(p_i)-Q_0(p_{i-1})d_0(p_i)\}^2}{p_i-p_{i-1}}, \qquad (2.6b)$$

$$K_3=\sum_{i=1}^{k+1}\frac{\{d_0(p_i)-d_0(p_{i-1})\}\{Q_0(p_i)d_0(p_i)-Q_0(p_{i-1})d_0(p_{i-1})\}}{p_i-p_{i-1}} \qquad (2.6c)$$

with $p_0=0$, $p_{k+1}=1$, and $d_0(p_0)=d_0(p_{k+1})=0$ and $n_i=[np_i]+1$, $i=1,2,\dots,k$.

The variance of $\hat{Q}(\xi)$ is given by

$$Var\left(\hat{Q}(\xi)\right)=\frac{\sigma^2}{n\Delta}\{K_2+Q_0^2(\xi)K_1-2Q_0(\xi)K_3\}. \qquad (2.7)$$

If the pdf, f_0, is symmetric about zero and if we select symmetric sample quantiles, i.e. $p_i+p_{k-i+1}=1$, $i=1,2,\dots,k$, then $K_3=0$ and the ABLUE of $Q(\xi)$ is given by

$$\hat{Q}(\xi)=X/K_1+Q_0(\xi)Y/K_2 \qquad (2.8)$$

with variance $Var\left(\hat{Q}(\xi)\right)=\frac{\sigma^2}{n}\left(1/K_1+Q_0^2(\xi)/K_2\right)$. (2.9)

These results are due to Ogawa (1951).

In order to obtain the optimal estimator of $Q(\xi)$, say $\hat{Q}^0(\xi)$, based on optimally chosen order statistics we minimize (2.7) with respect to $(p_1,p_2,\dots,p_k)$ subject to $0<p_1<p_2<\dots<p_k<1$. Let the optimum spacing be $(p_1^0,p_2^0,\dots,p_k^0)$, then the optimum ranks of the order statistics are $n_i^0=[np_i^0]+1$, $i=1,2,\dots,k$. The coefficients may also be computed based on $p_1^0,p_2^0,\dots,p_k^0$ by following through the formulas (2.4) and (2.5).

3. ABLUE of $Q(\xi)$ for the exponential distribution.

Let $X_{(n_1)}<X_{(n_2)}<\dots<X_{(n_k)}$ be k arbitrary but fixed order statistics with ranks $n_i=[np_i]+1$, $i=1,2,\dots,k$ in a sample of size n from the exponential distribution

$$F(x)=1-\exp\left\{-\frac{x-\mu}{\sigma}\right\}, \qquad x\ge\mu,\ \sigma>0, \qquad (3.1)$$

where μ, σ are unknown parameters. Then, using the results of Saleh and Ali (1966), and the results of Section 2, the ABLUE of $Q(\xi)$ is given by

$$\hat{Q}(\xi)=\{1+b_1(Q_0(\xi)-Q_0(p_1))\}X_{(n_1)}+\sum_{i=2}^{k}(b_i(\xi)-Q_0(p_1))X_{(n_i)} \qquad (3.2)$$

where $$b_1=-\frac{\{Q_0(p_2)-Q_0(p_1)\}}{\{e^{Q_0(p_2)}-e^{Q_0(p_1)}\}L}, \qquad (3.3a)$$

$$b_i=\frac{1}{L}\left\{\frac{Q_0(p_i)-Q_0(p_{i-1})}{e^{Q_0(p_i)}-e^{Q_0(p_{i-1})}}-\frac{Q_0(p_{i+1})-Q_0(p_i)}{e^{Q_0(p_{i+1})}-e^{Q_0(p_i)}}\right\}, \qquad (3.3b)$$

$i=2,\dots,k-1$,

$$b_k=\frac{1}{L}\left\{\frac{Q_0(p_k)-Q_0(p_{k-1})}{e^{Q_0(p_k)}-e^{Q_0(p_{k-1})}}\right\}, \qquad (3.3c)$$

and

$$L=\sum_{i=2}^{k}\frac{[Q_0(p_i)-Q_0(p_{i-1})]^2}{e^{Q_0(p_i)}-e^{Q_0(p_{i-1})}} \qquad (3.4)$$

The variance of $\hat{Q}(\xi)$ is given by

$$Var\left(\hat{Q}(\xi)\right)=\frac{\sigma^2}{n}\left\{\frac{[Q_0(\xi)-Q_0(p_1)]}{L}+e^{Q_0(p_1)}-1\right\}. \qquad (3.5)$$

In order to obtain the ABLUE based on k selected order statistics, we minimize (3.5) subject to the restriction, $0<p_1<p_2<\dots<p_k<1$. For this, we consider the alternative form

$$Var\left(\hat{Q}(\xi)\right)=\frac{\sigma^2}{n}\left\{\frac{t^2e^{Q_0(\xi)-t}}{K_2^{(k-1)}}+e^{Q_0(\xi)-t}-1\right\} \qquad (3.6)$$

where $t=Q_0(\xi)-Q_0(p_1)$ and $t_i=Q_0(p_{i+1})-Q_0(p_1)$, $i=1,2,\dots,k-1$,

and $$K_2^{(k-1)}=\sum_{i=1}^{k-1}\frac{(t_i-t_{i-1})^2}{e^{t_i}-e^{t_{i-1}}} \qquad \text{with } t_0=0. \qquad (3.7)$$

In order to minimize (3.6), we first maximize $K_2^{(k-1)}$ with respect to $t_1,\dots,t_{k-1}$. The maximum value is, say, K_2^0 occurring at $\left(t_1^0,\dots,t_{k-1}^0\right)$. Theorems relating to the maximization of $K_2^{(k-1)}$ are given in Saleh and Ali (1966) with tabulated values in Sarhan and Greenberg (1962). Next we minimize

$$t^2\exp\{Q_0(\xi)-t\}/k_2^0+\exp\{Q_0(\xi)-t\}-1 \qquad (3.8)$$

over the region $t\le Q_0(\xi)$, since $t\le Q_0\left(Q_0(p_1)\right)$ and $Q_0(p_1)$ is nonnegative. By differentiation, one finds that (3.8) is decreasing over $\left(0,Q_0(\xi_{K_2})\right]\cup\left[2-Q_0(\xi_{K_2}),\infty\right)$ and increasing over $\left(Q_0(\xi_{K_2}),2-Q_0(\xi_{K_2})\right)$ where $Q_0(\xi_{K_2})=1-(1-K_2^0)^{1/2}$. Now, we must have $t\le Q_0(\xi)$, so if $Q_0(\xi)\ge Q_0(\xi_{K_2})$, we get (3.4) equals

$$\left(Q_0^2(\xi)/K_2^0+1\right)\exp\left(Q_0(\xi)-Q_0(\xi_{K_2})\right)-1 \qquad (3.9)$$

at $t=Q_0(\xi_{K_2})$ and $Q_0^2(\xi)/K_2^0$ (3.10)

at $t=Q_0(\xi)$. These two values are equal at $t=Q_0(\xi)$ and $t=Q_0(\xi_{K_2}^*)$ where $t=Q_0(\xi_{K_2}^*)$ is located beyond $2-Q_0(\xi_{K_2})$ and is obtained by solving the equation $(3.9)-(3.10)$. Hence, we get the minimum of (3.8) at $t=Q_0(\xi)$, $Q_0(\xi)\in\left(Q_0(\xi_{K_2}),Q_0(\xi_{K_2}^*)\right]$, and $K_2^{(k-1)}=K_2^0$. Thus, when $\xi\in\left(\xi_{K_2},\xi_{K_2}^*\right]$, we must have

$$Q_0(p_1^0)=Q_0(\xi)-Q_0(\xi_{K_2}) \qquad (3.11)$$
$$Q_0(p_{i+1}^0)=Q_0(\xi)+t_i^0-Q_0(\xi_{K_2}),\qquad i=1,\dots,k-1.$$

Hence, the optimum spacing of the ABLUE of $Q(\xi)$ is given by

$$p_1^0=1-(1-\xi)/(1-\xi_{K_2}) \qquad (3.12)$$
$$p_{i+1}^0=1-(1-\xi)(1-\lambda_i^0)/(1-\xi_{K_2}),\qquad i=1,\dots,k-1,$$

where $\lambda_i^0=1-\exp(-t_i^0)$, $i=1,\dots,k-1$. The ABLUE based on this spacing is $\hat{Q}^0(\xi)=c_0^0X_{[np_1^0]+1}+\sum_{i=1}^{k-1}c_i^0X_{[np_{i+1}^0]+1}$ (3.13)

where $c_0^0=1+b_0^0Q_0(\xi_{K_2})$, $c_i^0=b_i^0Q_0(\xi_{K_2})$, $i=1,\dots,k-1$.

These coefficients may be computed easily using the tabulated values of $b_0^0,\dots,b_{k-1}^0$ and $\lambda_1^0,\dots,\lambda_{k-1}^0$ from Sarhan and Greenberg (1962).

For the case $\xi\notin\left(\xi_{K_2},\xi_{K_2}^*\right]$, the infimum occurs at $t=Q_0(\xi)$ and (2.8) is decreasing over $\left(0,\xi_{K_2}\right)\cup\left(\xi_{K_2}^*,1\right)$. Thus $Q_0(p_1)$ must be as small as possible and by using the results of Saleh and Ali (1966) we choose

$$Q_0(p_1^0)=\ln\left(\frac{2n-1}{2n+1}\right)^{-1},\ Q_0(p_{i+1}^0)=t_i^0,\ i=1,\dots,k-1. \qquad (3.14)$$

For this case, the ABLUE of $Q(\xi)$ is given by

$$\hat{Q}^0(p_1^0)=\left\{1+b_0^0\left(Q_0(\xi)+\ln\frac{2n-1}{2n+1}\right)\right\}X_{(1)}$$
$$+\sum_{i=1}^{k}b_0^0\left(Q_0(\xi)+\ln\frac{2n-1}{2n+1}\right)X_{(n_i^0)} \qquad (3.15)$$

where $b_1^0,\dots,b_{k-1}^0$ are tabulated in Table II.D.1 of Sarhan and Greenberg (1962) with $b_0^0=-\sum_{i=1}^{k-1}b_1^0$ and spacing $\left(p_1^0,p_2^0,\dots,p_k^0\right)$ defined by

$$p_i^0=\frac{1}{n+1/2},\ p_{i+1}^0=\left\{2+(2n-1)\lambda_i^0\right\}/(2n+1),\ i=1,\dots,k-1, \qquad (3.16)$$

where $\lambda_1^0,\dots,\lambda_{k-1}^0$ are spacings which maximize $K_2^{(k-1)}$ are tabulated in Sarhan and Greenberg (1962). Thus, the ABLUE of $Q(\xi)$ is given by (3.13) for $\xi\in\left(\xi_{K_2},\xi_{K_2}^*\right]$ and by (3.15) for $\xi\in\left(0,\xi_{K_2}\right]\cup\left(\xi_{K_2}^*,1\right)$. It may be noted that the estimate of $Q(\xi)$ depends on where ξ is located in the interval $(0,1)$.

Now, the complete sample BLUE is given by

$$\bar{Q}(\xi)=\frac{n}{n-1}(1-Q_0(\xi))X_{(1)}-\frac{1}{n-1}(1-nQ_0(\xi))\bar{X} \qquad (3.17)$$

$Q(\xi)$ with variance

$$Var\left(Q(\xi)\right)=\frac{\sigma^2}{n(n-1)}\left(1-2Q_0(\xi)+nQ_0^2(\xi)\right). \qquad (3.18)$$

The $ARE(\hat{Q}^0(\xi):\bar{Q}(\xi))$ is then given by

$$ARE(\hat{Q}(\xi):\bar{Q}(\xi))=\begin{cases} K_2^0 & ,\xi\in(0,\xi_{K_2}]\cup(\xi^*_{K_2},1) \\ Q_0^2(\xi)\left[\dfrac{\left\{\dfrac{Q_0^2(\xi_{K_2})}{K_2^0}+1\right\}}{\exp\{Q_0(\xi)-Q(\xi_{K_2})\}-1}\right]^{-1} & ,\xi\in(\xi_{K_2},\xi^*_{K_2}] \end{cases} \quad (3.19)$$

It may be noted that $\lim_{K\to\infty} K_2^* = 1$, hence $\lim_{K\to\infty} Q_0(\xi_{K_2}) = 1$, i.e. $\xi_1 = 1-e^{-1} = 0.6329 = \xi_1^*$ implies $Q_0(\xi_1^*) = 1$ and the spacing are the same as that of (3.14) and $ARE(\hat{Q}^0(\xi):\bar{Q}(\xi)) = 1$. Therefore, for large values of k we always use (3.15) and for small values of k we use $\hat{Q}^0(\xi)$. Table 1 provides some ARE-values.

Table 3.1

$ARE(\hat{Q}^0(\xi):\bar{Q}(\xi))$

k / ξ	2	4	6	8	10	12	14
0.60	0.7705	0.9095	0.9512	0.9697	0.9894	0.9848	0.9882
0.70	0.8115	0.9275	0.9606	0.9752	0.9894	0.9869	0.9869
0.80	0.8147	0.9108					
0.90	0.7205	0.9108					
K_2^0	0.6475	0.8910	0.9476	0.9693	0.9798	0.9857	0.9894
ξ_{K_2}	0.3339	0.4882	0.5375	0.5617	0.5759	0.5814	0.5912
$\xi^*_{K_2}$	0.9296	0.8291	0.7779	0.7476	0.7278	0.7137	0.7931

Note that ξ_{K_2} and $\xi^*_{K_2}$ are bounds on ξ for each k and K_2^0 are the ARE values for ξ outside the bounds.

3.1 Examples

Case 1: *Uncensored case.* Let $n=50$ and $k=2$. Then, from Table 1, $\xi_{K_2} = 0.3339$. The optimum spacings are $p_1^0 = 2/101 = 0.0198$,

$p_2^0 = (2+99\lambda_1^0)/(101) = (2+99(0.7968))/(101) = 0.8008$

when $\xi\in(0,0.3339]\cup(0.9296,1)$. The optimum ranks of the order statistics are 1 and 41. The coefficients for the estimate are

$c_0^0 = 1-(0.6275)(Q_0(\xi)-0.02) = 0.9874-0.6275Q_0(\xi)$

Ali et al. (1981) have $\quad c_0^0 = 1-Q_0(\xi)/1.5936$,

$c_1^0 = 0.6275(Q_0(\xi)-0.02) = 0.6275Q_0(\xi)-0.0126$.

Ali et al. (1981) have $\quad c_0^0 = Q_0(\xi)/1.5936$.

The difference is due to the fact that they use 1 in place of $(2n-1)/(2n+1)$.

$$ARE(\hat{Q}^0(\xi):\bar{Q}(\xi))=\begin{cases} 0.6476 & ,\xi\in(0,0.3339]\cup(0.9296,1) \\ \ln^2(1-\xi)/\left[0.8359(1-\xi)^{-1}-1\right] & ,\xi\in(0.3339,0.9296] \end{cases}$$

Case 2: *Censored case.* Let $n=50$ and $k=2$ with $p_i = 0.10$. Then, using Table 1 again, $\xi_{K_2} = 0.3339$. The optimum spacings when $\xi\in(0,0.4005]\cup(0.9366,1)$ are

$p_1^0 = 0.10$ and $p_2^0 = 0.10+(0.90)(0.7968) = 0.8888$.

The optimum ranks are $n_1^0 = 5$ and $n_2^0 = 45$. The coefficients are given by

$c_0^0 = 1-(0.6275)(Q_0(\xi)-0.1054) = 1.0661-0.6275Q_0(\xi)$,

$c_1^0 = 0.6275(Q_0(\xi)-0.1054) = 0.6275Q_0(\xi)-0.0661$.

Let $\xi\in(0.4005,0.9366]$, then

$p_1^0 = 1-1.5013(1-\xi) = 1.5013\xi-0.5013$

$p_2^0 = 1-0.2032(1.5013(1-\xi)) = 1-0.30506(1-\xi) = 0.30506\xi-0.69494$.

4. Estimation of $Q(\xi)$ for the Logistic Distribution.

Let $X_{(n_1)} < X_{(n_2)} < ..... < X_{(n_n)}$ be the sample order statistics and let the sample quantile function, $Q(\xi_i)$, be defined by

$$Q(\xi_i) = X_{(i_n)}, \quad 0 < u_i < 1, \quad i = [n\xi_i] + 1. \tag{4.1}$$

Given a spacing, $T = \{\xi_1, ..., \xi_k\}$, (k real numbers such that $0 < \xi_1 < \xi_2 < ... < \xi_k < 1$) the corresponding sample quantile, $Q(\xi_1), ..., Q(\xi_k)$, have been shown by Mosteller (1946) to have a normal limiting distribution. This form of distribution has been used to develop formulae for the asymptotically best linear unbiased estimator, $\mu^*(T)$ and $\sigma^*(T)$, of μ and σ and their corresponding asymptotic relative efficiencies (see Gupta and Gnanadesikan (1966) and Ogawa (1951)) and since the estimators and their ARE's are a function of the spacing for the sample quantiles, to obtain optimal estimators an appropriate choice of T must be made. Thus, an optimal spacing is defined as a spacing which minimizes the ARE of $\mu^*(T)$, $\sigma^*(T)$, or $(\mu^*(T), \sigma^*(T))$.

4.1 Estimator of the Scale Parameter. Let c_i be the asymptotically best optimal spacings and b_i be the corresponding coefficients of the best linear unbiased estimator $\hat{\sigma}_k$. Let the spacing c_i be defined in relation to the rank i of order statistics $X_{i:n}$ and sample size as

$$i = [nc_i + 0.5], \; 0 < c_i < \tfrac{1}{2} \tag{4.2}$$

and since $\lim_{n \to \infty} \frac{i}{n} = c_i$, it can therefore be easily shown that the estimator of σ, $\hat{\sigma}_k$, is asymptotically unbiased.

Let R_i denote the i-th sample quasi-range

$$R_i = X_{(n-i)} - X_{(i+1)}. \tag{4.3}$$

Then based on work of Weke (2001) and equation (4.3), we propose to use

$$\hat{\sigma}_k = \frac{\sum_{i=1}^{k} b_i [R_i + R_{i-1}]}{2\tau \sum_{i=1}^{k} b_i E_{i,i+1:n}}, \qquad \tau = \frac{\pi}{\sqrt{3}} \tag{4.4}$$

where

$$E_{i,i+1:n} = E\left[X_{(i)} + X_{(i+1)}\right] = Q(c_i) + Q'(c_i)\left(\frac{i}{n} - c_i\right)$$
$$+ \frac{Q''(c(i))}{2n^2}\left\{\frac{1}{6} + (i - nc_i)^2 - \frac{1}{4}\left(\frac{1}{\Delta_i} + \frac{1}{\Delta_{i+1}}\right)\right\} + O(n^{-3}) \tag{4.5}$$

as the estimator of the scale parameter σ. Since logistic distribution is symmetric about the mean, it follows that

$$E_{i,i+1:n} = -E_{n-i-1,n-i:n}, \qquad i = 1,2,...,n-1.$$

$E_{i,i+1:n}$ leads to the following expression when expanded around $c_i^* = \frac{i}{n}$ and $i = [nc_i + 0.5]$

$$E_{i,i+1:n} \approx \log\left(\frac{n-i}{i}\right) + \frac{n(n-2i)}{2i^2(n-i)^2}\left[\frac{1}{6} - \frac{2i}{(2i-1)(2i+1)}\right]. \tag{4.6}$$

The precision of this estimator is measured by calculating its bias and variance. The bias of the estimator $\hat{\sigma}_1$ is measured by

$$BIAS = 1 - E\left(\frac{\hat{\sigma}_k}{\sigma}\right) = 1 - \frac{\sum_{i=1}^{k} b_i E[R_i + R_{i-1}]}{2\tau \sum_{i=1}^{k} b_i E_{i,i+1:n}}. \tag{4.7}$$

Based on the values of the bias for various spacing values and sample sizes (see Tables 2 and 3) we notice that the bias values are negligible. Thus, the denominator in Equation (4.4) reasonably approximates the expectation of the selected order statistics. Hence, the variance of the estimator $\hat{\sigma}_k$ is given by

$$Var(\hat{\sigma}_k) = \frac{\sum_{i=1}^{k}\sum_{j=1}^{k} b_i b_j Cov(R_i + R_{i-1}, R_j + R_{j-1})}{\left[\sum_{i=1}^{k} b_i E(R_i + R_{i-1})\right]^2} \tag{4.8}$$

where the ranks i and j are determined from Equation (4.2) and b_i and b_j are the asymptotic optimal coefficients corresponding to the ranks i and j.

The expectations, variances and covariances of the logistic order statistics for sample sizes up to 100 were computed by using FORTRAN 77 language computer programs to facilitate the calculations of the estimator and its derived properties.

The expectation (EXP) of the selected order statistics, bias (BIAS) and variances (VAR) of the estimator for various spacing values and sample sizes are given in Tables 2 - 3. The asymptotic efficiency (AEF) of the estimator with respect to the Cramer-Rao lower variance bound and the relative efficiency (R.E.) with respect to the variance of the best linear unbiased estimator are computed and presented in Tables 2 - 3 for comparison. Notice that

$$AEF = \frac{9}{n(3+\pi^2)Var(\hat{\sigma}_k)}$$

and

$$R.E. = \frac{Var(BLUE(\sigma))}{Var(\hat{\sigma}_k)}. \tag{4.9}$$

The asymptotically optimum spacings, coefficients, the variances and the asymptotic relative efficiencies of the estimator of scale parameter for the logistic distribution were first provided by Hassanein (1969) for $2 \le k \le 9$. Later, Eubank (1981) worked out asymptotically best optimal spacings, coefficients and the corresponding asymptotic relative efficiencies for logistic distribution for $2 \le k \le 10$. Eubank's optimal spacings generated higher relative efficiencies than the efficiencies in Hassanein (1969). In this study, the asymptotically best optimal spacings and coefficients in Eubank (1981) have been used for $k = 4,6,8$ to produce relatively higher asymptotic efficiencies.

Table 4.1

EXP, BIAS, VAR, AEF and R.E. of $\hat{\sigma}_2$

SIZE	EXP	BIAS	VAR	AEF	R.E.
14	13.12119	-.00001	.05613	.88999	.9507
15	13.77081	-.00001	.05192	.89795	.9549
16	12.29594	-.00001	.04920	.88839	.9409
17	12.88540	-.00001	.04578	.89858	.9484
18	13.43297	-.00001	.04287	.90635	.9536
19	13.94430	-.00001	.04035	.91210	.9571
20	14.42382	-.00001	.03817	.91607	.9589
21	13.28171	.00000	.03692	.90199	.9420
22	13.72777	.00000	.03503	.90750	.9460
23	14.14942	.00000	.03335	.91161	.9487
24	14.54927	.00000	.03186	.91460	.9498
25	13.62037	.00000	.03121	.89627	.9295
26	13.99673	.00000	.02987	.90046	.9324
27	14.35547	.00000	.02867	.90331	.9344
28	14.69835	.00000	.02758	.90560	.9355
29	15.02671	.00000	.02657	.90749	.9364
30	14.24064	.00000	.02625	.88813	.9154
31	14.55308	-.00001	.02531	.89132	.9174
32	14.85315	.00000	.02448	.89279	.9183
33	15.14215	-.00001	.02369	.89466	.9194
34	14.46276	.00000	.02353	.87398	.8976
35	14.73915	.00000	.02282	.87562	.8983
36	15.00611	-.00001	.02213	.87769	.9001
37	15.26399	.00000	.02152	.87836	.9001
38	15.51353	.00000	.02095	.87863	.8993
39	14.91389	-.00001	.02085	.85984	.8801
40	15.15406	.00000	.02032	.86046	.8794
41	15.38690	.00001	.01982	.86060	.8799
	∞	.00000	.85132/n	.82145	

Note: Optimal spacings are .036, .221
Coefficients: .1401, .3544

4.2 Numerical Examples and Discussions In this section, we use the k asymptotically optimal spacings and the k corresponding asymptotic optimal coefficients ($k = 2,3,4$) to construct Tables 2 - 3. A numerical example is given below for comparison and illustration:

Let $p(i) = [nc_i + 0.5]$, $i = 1,2,\ldots,k$ be the rank of order statistics. In this example, we consider the case when $n = 20$ and $k = 4$ so that the ranks are given as $p(1) = [0.036n + 0.5]$ and $p(2) = [0.221n + 0.5]$. Hence, the range of sample size applicable in this particular case is $14 \le n \le 41$ and by using the optimal coefficients as 0.1401 and 0.3544 given by Eubank (1981) the following computations ensues:

Let the coefficients be in the ratio of 1: 2.5296 and by using Equation (4.2) the order statistics are $i = 1,2,4,5$ and $n-i+1 = 16,17,19,20$ such that the calculations become

$i = 1$ $\qquad\qquad$ $i = 2$

$i = 4$

$$2\times\left\{\begin{array}{l} 1\times 0.51558 \\ 1\times 0.20390 \\ 2.5296\times\left[\begin{array}{l}0.09023\\0.07045\end{array}\right. \\ -2.5296\times\left[\begin{array}{l}0.02060\\0.01935\end{array}\right. \\ -1\times 0.01727 \\ -1\times 0.01638 \end{array}\right. \qquad 2\times\left\{\begin{array}{l} 1\times 0.21246 \\ 2.5296\times\left[\begin{array}{l}0.09457\\0.07393\end{array}\right. \\ -2.5296\times\left[\begin{array}{l}0.02170\\0.02039\end{array}\right. \\ -1\times 0.01820 \end{array}\right.$$

$$2\times\left\{\begin{array}{l} 2.5296\times 0.10469 \\ 2.5296\times\left[\begin{array}{l}0.08208\\-0.02431\end{array}\right. \\ -2.5296\times 0.10469 \end{array}\right.$$

$i = 5$

$$2.5296\times\left\{\begin{array}{l} 0.08688 \\ -0.02586 \end{array}\right.$$

$SUM_1 = 1.48358$, $SUM_2 = 0.833794$, $SUM_4 = 2.5296\times 1.263010$

$SUM_5 = 2.5196\times 0.154356$

$$SUM = SUM_1 + SUM_2 + SUM_4 + SUM_5 = 3.970523$$

$$EXP = 2\{1.95597 + 1.37563 + 2.5296(0.85312 + 0.68083)\} = 14.423760,$$

$$VAR = \frac{2\times SUM}{(EXP)^2} = 0.038170,$$

$$AEF = \frac{9}{20(3+\pi)^2 \times VAR} = 0.91606,$$

$Vopt = Var(BLUE) = 0.0366$ and so $R.E. = \dfrac{Vopt}{VAR} = 0.95887$.

Table 4.2

EXP, BIAS, VAR, AEF and R.E. of $\hat{\sigma}_3$

SIZE	EXP	BIAS	VAR	AEF	R.E.
34	38.50034	.00000	.02213	.92934	.9544
35	39.25244	.00000	.02146	.93120	.9553
36	38.44604	.00000	.02098	.92586	.9495
37	39.16250	.00000	.02038	.92721	.9504
38	39.85580	-.00001	.01980	.92929	.9515
39	37.84077	-.00001	.01923	.93249	.9542
40	37.11419	-.00002	.01875	.93254	.9531
41	37.76245	-.00001	.01826	.93403	.9551
42	38.39204	-.00002	.01777	.93718	.9567
43	37.73132	-.00001	.01743	.93318	.9530
44	38.33611	-.00003	.01696	.93696	.9558
45	38.92271	-.00002	.01659	.93660	.9554
46	39.49372	-.00001	.01622	.93720	.9544
47	38.88341	-.00002	.01593	.93408	.9517
48	39.43372	-.00002	.01559	.93437	.9500
49	39.96975	-.00001	.01528	.93416	.9503
50	38.41880	.00000	.01494	.93597	.9505
51	37.85509	-.00001	.01466	.93534	
52	38.36438	-.00001	.01437	.93603	
53	38.86168	.00000	.01408	.93698	
54	39.34752	.00000	.01381	.93765	
55	38.82023	-.00001	.01359	.93534	
56	39.29089	.00000	.01335	.93576	
57	39.75121	.00000	.01311	.93580	
58	39.25986	.00000	.01293	.93263	
59	39.70681	.00000	.01270	.93339	
60	40.14429	.00000	.01249	.93334	
64	39.25223	.00000	.01169	.93457	
65	39.65557	.00000	.01151	.93511	
68	39.99800	.00000	.01102	.93286	
70	39.96281	.00000	.01074	.92978	
72	39.26779	.00000	.01044	.93058	
75	39.58537	.00000	.01003	.92972	
80	40.55407	.00001	.00942	.92804	
84	39.26011	.00000	.00902	.92268	
85	39.57013	.00000	.00891	.92315	
88	39.85278	.00000	.00862	.92138	
90	40.43928	.00000	.00843	.92224	
95	40.16094	.00001	.00801	.91901	
96	39.86174	.00001	.00796	.91568	
100	40.37524	.00001	.00765	.91386	
∞		.00000	.78124/n	.89514	

Note: Optimal spacings are .015, .091, .268

Coefficients: .0509, .1938, .2468

These values correspond to the values of expectation, variance, asymptot efficiency and relative efficiency for $\hat{\sigma}_2$ given in Table 2.

It is important to note that the values of the variances in Table 2 are lower th those in Weke (2001) which were constructed by using a single spacing ar those given by Chan et al. (1971). And, therefore, the asymptotic efficienci and relative efficiencies are much higher.

Table 3 presents the expectation, bias, variance, asymptotic efficiency ar relative efficiency of the estimator $\hat{\sigma}_3$ for various n values. The table constructed by using three pairs of optimal spacings and it is noticed th increasing the number of spacings results in an increase in the efficiencies.

REFERENCES

Ali, M.M., D. Umbach and K.M. Hassanein (1981). Estimation of quantiles exponential and double-exponential distributions based on two ord statistics. *Commun. Statist. A-Theory/Methods* **10**, 1921-1932.

Balakrishnan, N. and N. Kannan (2001). Point and interval estimation of logist parameters based on Type II progressively censored samples. Handbook of Statistics – Vol. 20, North-Holland, Amsterdam.

Chan, L.K., N.N. Chan and E.R. Mead (1971). Best linear unbiased estimates the parameters of the logistic distribution based on selected ord statistics. *J.A.S.A.* **66**, 889-892.

Dixon, W.J. (1957). Estimates of the mean and standard of a normal populatio *Ann. Math. Statist.* **28**, 806-809.

Dixon, W.J. (1960). Simplified estimation from censored normal samples. *An Math. Statist.* **31**, 385-391.

Epstein, B. (1960). Estimation of life test data. *Technometrics* **2**, 447-454.

Eubank, R.L. (1981). A density-quantile function approach to optimal spacir selection. *Ann. Statist.* **9**, 494-500.

Gupta, S.S. and M. Gnanadesikan (1966). Estimation of the parameters of tl logistic distribution. *Biometrika* **53**, 565-570.

Hassanein, K.M. (1968). An Analysis of extreme value data by sample quantil for very large samples. *J. Amer. Statist. Assoc.* **63**, 877-888.

Hassanein, K.M. (1969). Estimation of parameters of the logistic distribution of sample quantiles. *Biometrika* **56**, 684-687.

Hassanein, K.M. (1972). Simultaneous estimation of the parameters of the extreme value distribution by sample quantiles. *Technometrics* **14**, 63-70.

Kubat, P. and B. Epstein (1980). Estimation of quantiles of location-scale distributions based on two or three order statistics. *Technometrics* **22**, 575-582.

Mann, N.R. (1970). Estimators and exact confidence bounds for Weibull parameters based on few ordered observations. *Technometrics* **12**, 345-361.

Mann, N.R. and K.W. Fertig (1977). Efficient unbiased quantile estimators for moderate size complete samples from extreme value and Weibull distributions. *Technometrics* **19**, 87-93.

Mosteller, F. (1946). On some useful 'inefficient' statistics. *Ann. Math. Statist.* **17**, 377-408.

Ogawa, J. (1951). Contributions to the theory of systematic statistics I. *Osaka Math. Journal* **3**, 175-213.

Parzen, E. (1979). Nonparametric statistical data modeling. *J. Amer. Statist. Assoc.* **74**, 105-121.

Saleh, A.K.Md.E., M.M. Ali (1966). Asymptotic optimum quantiles for the estimation of the negative exponential distribution. *Ann. Math. Statist.*, **37**, 143-151.

Sarhan, A.E. and B.G. Greenberg (1962). Simplified estimates for the exponential distribution. *Ann. Math. Statist.*, **34**, 102-116.

Weke, P.G.O. (2001). *Simplified Optimal and Relay Linear Unbiased Estimation of Parameters of Logistic Distribution*; Ph.D. Thesis - Harbin Institute of Technology, Department of Mathematics, July 2001.

The East African Journal of Statistics ISSN: 1811-7503
Volume1, Number 1, pp.109-116 ©2005 JKUAT Press Ltd

ON THE NON-LINEAR STOCHASTIC PRICE ADJUSTMENT OF SECURITIES

SILAS N. ONYANGO

Faculty of Science,
Department of Mathematics and Applied Statistics,
Maseno University,
P.O. Box 333,
Maseno, KENYA
Email: *s_onyango@hotmail.com*

Abstract. We introduce a type of Itô process that models the adjustment of the market price of a traded security to new information affecting supply and demand. It is based on supply and demand functions and the Walrasian price adjustment assumption that proportional price increase is driven by excess demand. When supply and demand curves are linearised about the equilibrium point, the process turns out to be a logistic form of Brownian motion with random element of the Wiener type.

Key words. Excess demand function, Walrasian price adjustment process, Logistic Brownian motion.

1. **Introduction.** The stochastic evolution of security prices freely traded in markets has been an abiding topic of research- since the publication of Louis Bachelier's dissertation –Théorie de la Speculation, Bachelier (1900). After the introduction of geometric Brownian motion models, Samuelson (1965), Black and Scholes, (1973), Merton (1973), the broad thrust of recent work has focused on steady market conditions defined by the well - known Itô process

$$dP_t = \mu P_t dt + \sigma P_t dZ \tag{1.1}$$

The linear drift coefficient μ reflects price trading driven by constant investor expectation of gain, while the diffusive Wiener process σdZ reflects the response of trading to random fluctuations in supply and demand.
In this paper we focus not on volatility but on the possibility of non-linear drifting as markets adjust to radically new perceptions of the price of security. The basic driver of such price adjustment, taking a neo-Walrasian view, Walras (1874), Samuelson (1941, 1947) is the excess demand over supply at the trading point. We use a linearised version of this driving force – *la puissance motrice de la spéculation*- to drive an Itô process that models price adjustment in non-steady markets. It is a logistic process with diffusive Wiener variation- 'logistic Brownian motion'.

2. **General Walrasian_Samuelson Price Adjustment Model.** Since the 1960s, a model of Walrasian, Walras (1874) *tatonnement* has been used to study stability of general price equilibrium, Samuelson (19650, Anderson (2000) and Asparouhova, Bossaerts & Plott, (2000) among others. In this paper, we take the core principle of the standard Walrasian model. That is: security price changes are directly driven by the excess demand for the security. For simplicity we do