*Journal of Statistical Research* *ISSN 0256 - 422 X*
*2006, Vol. 40, No. 1, pp. 1-21*
*Bangladesh*

# SMALL-SAMPLE PROPERTIES OF SOME IMPROVED ESTIMATORS IN LOGISTIC REGRESSION MODEL WITH SKEW-NORMALLY DITRIBUTED EXPLANATORY VARIABLES.

M. A. MATIN

*Department of Statistics, School of Mathematics*
*University of Leeds, Leeds, LS2 9JT, UK*

*Email: matin@maths.leeds.ac.uk*

A. K. MD. EHSANES SALEH

*School of Mathematics and Statistics*
*Carleton University, Ottawa, K1S 5B6, Canada*

*Email: esaleh@math.carleton.ca*

SUMMARY

This study explores the small-sample properties of five estimators (the unrestricted maximum likelihood estimator, the shrinkage restricted estimator, the shrinkage preliminary test estimator, the shrinkage estimator and the positive-rule shrinkage estimator) using Monte Carlo experiments to confirm the asymptotic findings of Matin and Saleh (2005). It also explores the properties of test procedures (the Wald, the score and the likelihood ratio) in performing in estimators and tests under consideration. This study confirms the theoretical results in cases where comparisons are possible. When the number of explanatory variables is greater than or equal to 3 the shrinkage and the positive-rule shrinkage estimators *always* perform well. Considering the MSE the positive-rule shrinkage estimator performs better than the shrinkage estimator. The likelihood ratio test stands out to be the best. However, we lean toward the use of the Wald statistic when the problem of estimation is of paramount interest as it provides lower bias and MSE for the estimators.

*Keywords and phrases:* Preliminary test estimator, Shrinkage estimator, Stein-type estimator, Wald test, Score test, Likelihood ratio test, Coverage probability, Graph analysis, Simulation.

*AMS Classification:* 62Jxx.

# 1 Introduction

Matin and Saleh (2005) deals with the problem of estimating the parameters of logistic regression model when it is known from extraneous sources that the *uncertain prior information* in the form of the hypothesis $H_0 : \beta_0 = \ldots = \beta_{k-1} = \beta^0$ (pivot) may hold. They have proposed five estimators namely the unrestricted maximum likelihood estimator (UMLE), the shrinkage restricted estimator (SRE), the shrinkage preliminary test estimator (SPTE), the shrinkage estimator (SE) and the positive-rule shrinkage estimator ($SE^+$). The SE and $SE^+$ are the Stein-type estimators based on the preliminary test approach of Saleh and Sen (1987). Matin and Saleh (2005) have sorted out the preferences toward the application of five proposed estimators under local alternatives through the analysis of asymptotic mean square error (MSE) matrix and distributional risk. It reveals that when $k \geq 3$, we should use the SE or $SE^+$ and for $k \leq 2$ it is advisable to use the preliminary test estimator (PTE).

Matin and Saleh (2005) made use of the test statistic Wald (WALD) in their proposed estimators SPTE, SE and $SE^+$. However, one can use the score statistic (SCT) as well as the likelihood ratio statistic (LRT) as they are asymptotically equivalent. In practice, they are used to test the significance of the parameter vector $\beta = (\beta_0, \beta_1, \ldots, \beta_{k-1})^T$. They have optimal asymptotic properties however, the small-sample behavior is less well known (see Matin (2005) for some detail results).

This attempt is made to explore the small sample properties of the said estimators to confirm the asymptotic findings using Monte Carlo experiments with skew-normally distributed explanatory variables. This also explores the small-sample properties of test procedures as performing in estimators and tests under consideration. Seection 2 describes the design along with the computer program of the study. Results with discussion are given in section 3 and summary in section 4.

# 2 Design of the Study

The logistic regression model considered here consists of a dichotomous dependent variable and two (four) skew-normally (see Azzalini (1985) and Henze (1986) for details) distributed explanatory variables with an intercept term (i.e., $k = 3$ (5)). The salient features of the skew-normal distribution are mathematical tractability and strict inclusion of the normal density as a special case ($\lambda = 0$); where the shape parameter $\lambda$ controls the index of skewness. Two degrees of skewness designated by the values -7.0 and 0.0 of the shape parameter were taken into account to represent negative and no skewness (better feeling of the skewness implied by our choice of shape parameter can be seen in Matin (2005)) in the explanatory variables. As there is a great influence of sample size on the bias of regression coefficients, four different sample sizes 40, 50, 100 and 200 were considered. Two sets of regression coefficients (1, 1, 1) ((1, 1, 1, 1, 1)) and (2, 2, 2) ((2, 2, 2, 2, 2)) were included in the design for $k = 3$ (5). Three test statistics (WALD, SCT and LRT) were considered to test the null hypothesis ($H_0 : \beta_0 = \beta_1 = \beta_2 = 2$ (1) i.e., $\beta^0 = 2$ (1) for $k = 3$ and $H_0 : \beta_0 = \beta_1 = \beta_2 = \beta_3 = \beta_4 = 2$ (1) i.e., $\beta^0 = 2$ (1) for $k = 5$) and three different levels of significance ((.40, .45, .50) for $k = 3$ and (.10, .15, .20) for $k = 5$) were considered for each of the test statistic to see how the preliminary test estimator behaves with an increase or decrease of the significance level. One of the three significance levels for $k = 3$ (5) was chosen by the Akaike information criterion with $\alpha^* = P(\chi^2_k > 2)$ as the optimal level and the other two were chosen as neighborhood values to the optimal level. However, to make the design of the study less burdensome we excluded some parameter combinations. The number of replications was 1000 for each experiment.

To perform the Monte Carlo experiments, a FORTRAN program was written in double precision. The pseudo random numbers were generated by the Wichmann and Hill (1982) algorithm. The pseudo standard normal variate was generated by the Polar Marsaglia method (Morgan, 1984). The skew-normal distribution was generated by exploiting the result

$$Z := \frac{\lambda}{\sqrt{1+\lambda^2}}|U| + \frac{1}{\sqrt{1+\lambda^2}}V \sim SN(\lambda)$$

where $U$ and $V$ were independent standard normal random variables given by Henze (1986). Other routines needed to run the program were taken from the SLATEC library and from a book on algorithm edited by Griffith and Hill (1985). The Newton-Raphson method was used to maximize the log likelihood function. The condition for a Newton-Raphson iteration to converge was set to the absolute value of $(\widehat{\beta}_{(i+1)} - \widehat{\beta}_{(i)})/\widehat{\beta}_{(i)}$ less than or equal to $10^{-6}$. The maximum number of iterations was fixed at 10 which is sufficient enough for our purpose. The convergence of the iterative procedure used in general is very fast. Additional pre-simulation experiments indicated that the maximum iteration limit of 10 was a reasonable choice and did not lead to the "throwing away" of "good samples" (samples where convergence would have been obtained with a higher iteration maximum).

The dependent variable ($Y$) was generated from a uniform random variable ($u(0\ 1)$) conditional on $\mathbf{x}$ as in equation,

$$log\frac{\pi}{1-\pi} = \beta_0 + \beta_1 x_1 + \ldots + \beta_{k-1}x_{k-1} \qquad (2.1)$$

whereby $Y = 1$ with probability $[1 + e^{-\beta^T\mathbf{x}}]^{-1}$, and $= 0$ otherwise. The chosen parameter set was used to compute the probability mentioned above where $\mathbf{x}$ was generated as a skew-normal variate. Once the dependent variable and the matrix of explanatory variables were generated the program computed the regression estimates for the logistic model first by unrestricted ML method and then by other methods. It should be noted here that we have used the term $\frac{(k-2)(n-k)}{n-k+2}$, as a small sample adjustment, instead of $k - 2$ in the formulas for the estimators SE and SE$^+$ (see Matin and Saleh (2005)). Note that as $n \to \infty$, $\frac{(k-2)(n-k)}{n-k+2} \to k - 2$. The program also done the necessary computation regarding the tests WALD, SCT and LRT.

An obvious limitation of this study is that we only consider the case where the null hypothesis is true. However, small-scale investigations confirm the asymptotic results under the local alternatives too. To have a better understanding of the behavior of the estimators and tests under local alternatives further detailed simulation study is necessary.

# 3 Results and Discussion

## 3.1 Description of the Results

In order to study the properties of the estimators and their relative performance, the bias and mean square error were computed for each estimator, for each sample size and for each parameter. To have a general idea about overall performance of each of the estimators norm bias (NBIAS) and summed MSE (SMSE) were also computed (Table 1, Table 2 and Table 3). The norm of a vector $x \in R^n$ is given by $||x|| = (x.x)^{1/2} = (\sum x_i^2)^{1/2}$ while the SMSE of an estimator $\beta^*_{\mathbf{n}}$ is defined as $\sum\{E(\beta^*_{\mathbf{n}} - \beta)^{\mathbf{2}}\}$.

To judge the capability of the estimators to reflect the true parameter value, we report the coverage probability (CP) (Table 1, Table 2 and Table 3) of 95% confidence set for the estimators by using the formula

$$P_\beta\{\beta \in C_a(X)\} = P_\beta\{\beta : (\beta^*_n - \beta)^T (X^T V X)(\beta - \beta^*_n) \leq \chi^2_{k,\alpha}\} = 1 - \alpha \qquad (3.1)$$

where $\beta^*_n$ standing for an estimator and the formula defines the $1 - \alpha$ confidence ellipsoid for $\beta$. In doing so, we followed the George and Casella (1994) formulation of the recentered confidence set. *Recentered set estimators* are the *set estimators* where the usual "centre" has been replaced by an alternative estimator (see Robert and Saleh (1989)).

Given the observation of a $k$-dimensional multivariate normal vector $X \sim N_k(\beta, I)$ the classical confidence set for $\beta$ is defined by $C_0(X) = \{\beta : ||\beta - X||^2 \leq c^2\}$. For all $\beta$ the coverage probability, $P_\theta(\beta \in C_0(X)) = P\{\chi^2_k \leq c^2\} = 1 - \alpha$, is consistent and in practice we report that $C_0(X)$ contains $\beta$ with confidence $1 - \alpha$.

Hwang and Casella (1982, 84) showed that for $k \geq 3$ the set $C_0(X)$ can be improved by recentering at a positive-rule Stein estimator $\delta_a(X) = u_a(||X||)$, where $u_a(r) = max\{[1 - (a/r^2)],\ 0\}$ to obtain $C_a(X) = \{\beta : ||\beta - \delta_a(X)||^2 \leq c^2\}$. $C_a(X)$ dominates $C_0(X)$ in the sense that both sets possess equal volume but for $a$ in a certain range, $C_a(X)$ has uniformly higher coverage probability for all $\beta$, that is

$$P_\beta\{\beta \in C_a(X)\} > P_\beta\{\beta \in C_0(X)\} = 1 - \alpha. \qquad (3.2)$$

Thus, we are better off in reporting $C_a(X)$ than $C_0(X)$ as a $1 - \alpha$ confidence region for $\beta$.

These recentered confidence sets are merely pointing out the inadequacy of the usual confidence sets rather a useful alternative. Criticism of being related to the Neyman-Pearsonian approach is still hold. Their coverage probability is always above the nominal level $1 - \alpha$ while the minimum is still $1 - \alpha$ (Robert and Saleh, 1989).

To show the properties of the three large-sample test statistics, the mean, standard deviation, mean $p$-value, proportion significant (Table 4) and quantiles (Table 5) were computed for each of the three tests over the number of replications. In doing so, we considered only the case of $\lambda = 0$. The standard errors for the mean $p$-value and proportion significant were also computed by the square root of the quantity $\widehat{p}(1-\widehat{p})/(\text{number of replications})$. The $\alpha$-th quantile was computed by the formula

$$\begin{aligned} \widehat{v}_\alpha = V_{(1)} & \qquad \text{if } K = 1 \\ \gamma V_{(K-1)} + (1-\gamma)V_K & \qquad \text{if } 1 < K \leq m \\ V_{(m)} & \qquad \text{if } K = m+1 \end{aligned}$$

where $V_{(1)} < V_{(2)} < \ldots < V_{(m)}$ are the order statistics for $V_i$ $i = 1, 2, \ldots, m$. Here $0 < \alpha < 1$, $K = [(m+1)\alpha] + 1$ and $\gamma = \{K - (m+1)\alpha\}$. The motivation behind the use of this formula was provided by Lewis and Orav (1989) in page 149. In our case $\alpha \in \{.10, .20, .30, .40, .50, .60, .70, .80, .90, .95, .99\}$. To facilitate the comparison between the empirical and theoretical quantiles of the test statistics, we present the theoretical quantiles of a chi-square distribution for the degrees of freedom 3 (5) in Table 5.

## 3.2 Bias

In general, as the sample size increases the bias for each of the regression coefficient of UMLE, SRE($c = .50$) and SPTE ($c = .50$) decreases however, the exceptions are SPTE($c = 1.00$), SE and SE$^+$ particularly for small samples. In large samples, the biases are remarkably close to zero. Generally, for a specific sample size with $k = 3$ the biases are found to be smaller than the case with $k = 5$ for the corresponding regression coefficients. For the estimators where the level of significance $\alpha$ is directly involved, as $\alpha$ increases the bias of each regression coefficient also increases. Theoretically it is possible to have larger biases for the estimators compared to the UMLE however, we did not encounter such a possibility. In the worst possible case the estimators can reduce at least 10% of the bias incurred in the UMLE while the best possible scenario can reach up to 50%. Considering the observed biases, the estimators can be ordered as SPTE($c = 1.00$) < SRE ($c = .50$) < SPTE($c = .50$). No clear conclusion can be drawn about the SE and SE$^+$ however, the SE (SE$^+$) < SRE($c = .50$). For the choice of test statistics, we observed differences in the biases (of the regression coefficients) of SE, SE$^+$ with $k = 3$ (5) which follows a regular pattern $\text{Bias}_{LRT} > \text{Bias}_{SCT} > \text{Bias}_{WALD}$. Considering the use of different level of $\alpha$, the bias-differences are found to be smaller with the WALD however, larger with the SCT and LRT.

So far we have discussed the results for the case of $\beta^0 = 2$. Comparing this case (for $k = 5$ with the LRT) with the case of $\beta^0 = 1$, the biases are smaller in the latter case for a specific sample size (for the corresponding regression coefficients) than the former for all the estimators. (This gives an indication of how the bias changes with the changes in the size of the parameter.) In this case, as the sample size increases the biases decrease for all the estimators considered. In large samples, the biases are remarkably close to zero.

So far we have discussed above the results for the case of $\lambda = 0$. Considering a single case where $\lambda = -7$ (for $k = 5$ with the LRT) with the case of $\beta^0 = 1$, the biases are found to be larger than the case of its counterpart $\lambda = 0$.

The NBIASs of the estimators portray a better scenario of the remarks we have just made above. We can order the estimators according to their NBIASs more clearly.

## 3.3 Mean Square Error

In general, as the sample size increases the mean square error for each of the regression coefficients of all the estimator decreases. Generally, for a specific sample size with $k = 3$ the MSEs are found to be smaller than the case with $k = 5$ for the corresponding regression coefficient. For the estimators where the level of significance $\alpha$ is directly involved, as the level of $\alpha$ decreases the MSEs also decrease. Considering the observed MSEs, the estimators can be ordered as SPTE ($c = 1.00$) $<$ SRE($c = .50$) $<$ SPTE($c = .50$), SE$^+$ $<$ SE and SE (SE$^+$) $<$ SRE($C = .50$). For the choice of test statistics, the MSEs (of the regression coefficients) of SPTE, SE and SE$^+$ are found to be different with $k = 3$ (5) that follows a regular pattern MSE$_{LRT}$ $>$ MSE$_{SCT}$ $>$ MSE$_{WALD}$. Considering the different level of $\alpha$, the MSE-differences are found to be smaller when the WALD statistic is used however, larger with the SCT and LRT.

So far we have discussed the results for the case of $\beta^0 = 2$. Comparing this case (for $k = 5$ with the LRT) with the case of $\beta^0 = 1$, we observed that the MSEs are smaller in the latter case for a specific sample size (for the corresponding regression coefficients) than the former for all the estimators; the exceptions are the case of SPTE where the MSEs increase. In this case, as the sample size increases the MSEs decrease for each of the regression coefficient for all the estimators.

So far we have discussed above the results for the case $\lambda = 0$. Considering a single case where $\lambda = -7$ (for $k = 5$ with the LRT) with $\beta^0 = 1$, we found that the MSEs are extremely larger than its counterpart $\lambda = 0$ however, regarding the MSE reduction (to the UMLE) the scenario remains almost the same. For large samples, particularly for the sample size 200 the MSEs for all the estimators are remarkably close to each other.

The SMSEs of the estimators portray a better scenario of the remarks we have just made above. We can order the estimators according to their SMSEs more clearly.

## 3.4 Coverage Probability

In general, as the sample size increases the coverage probabilities of UMLE, SPTE($c = .50$) and SPTE($c = 1.00$) tend to .950. In addition, with $k = 3$ the CPs are remarkably close to .950 while with $k = 5$ they are slightly different. Generally, for a specific sample size with $k = 5$ the CPs are larger than the case of $k = 3$. Considering the SPTE where the level of significance $\alpha$ is directly involved, we did not find any significant difference in the CPs for different levels of $\alpha$, the only exception being the case with $k = 5$ when the LRT

Table 1: NBIAS, SMSE and CP of the estimates for $k = 3$ with $\lambda = 0$ and $\beta^0 = 2$.

| | | WALD | | | SCT | | | LRT | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | n | NBIAS | SMSE | CP | NBIAS | SMSE | CP | NBIAS | SMSE | CP |
| UMLE | 40 | 0.8895 | 5.7079 | 0.961 | 0.8895 | 5.7079 | 0.961 | 0.8895 | 5.7079 | 0.961 |
| | 50 | 0.7255 | 3.7808 | 0.959 | 0.7255 | 3.7808 | 0.959 | 0.7255 | 3.7808 | 0.959 |
| | 100 | 0.3540 | 1.3416 | 0.965 | 0.3540 | 1.3416 | 0.965 | 0.3540 | 1.3416 | 0.965 |
| | 200 | 0.1564 | 0.4513 | 0.954 | 0.1564 | 0.4513 | 0.954 | 0.1564 | 0.4513 | 0.954 |
| SRE(0.50) | 40 | 0.4447 | 1.4270 | 1.000 | 0.4447 | 1.4270 | 1.000 | 0.4447 | 1.4270 | 1.000 |
| | 50 | 0.3628 | 0.9452 | 1.000 | 0.3628 | 0.9452 | 1.000 | 0.3628 | 0.9452 | 1.000 |
| | 100 | 0.1770 | 0.3354 | 1.000 | 0.1770 | 0.3354 | 1.000 | 0.1770 | 0.3354 | 1.000 |
| | 200 | 0.0782 | 0.1128 | 1.000 | 0.0782 | 0.1128 | 1.000 | 0.0782 | 0.1128 | 1.000 |
| SPTE(0.50) | 40 | 0.5034 | 2.8536 | 0.961 | 0.5952 | 3.8555 | 0.961 | 0.7870 | 5.1213 | 0.961 |
| $\alpha = .40$ | 50 | 0.4741 | 2.2825 | 0.959 | 0.5447 | 2.9201 | 0.959 | 0.6477 | 3.3870 | 0.959 |
| | 100 | 0.2660 | 1.0752 | 0.965 | 0.3124 | 1.0945 | 0.965 | 0.3124 | 1.1678 | 0.965 |
| | 200 | 0.1198 | 0.3520 | 0.954 | 0.1198 | 0.3545 | 0.954 | 0.1365 | 0.3676 | 0.954 |
| SPTE(0.50) | 40 | 0.5296 | 3.0949 | 0.961 | 0.6320 | 4.1827 | 0.961 | 0.8005 | 5.2148 | 0.961 |
| $\alpha = .45$ | 50 | 0.5065 | 2.5541 | 0.959 | 0.5713 | 3.0905 | 0.959 | 0.6570 | 3.4356 | 0.959 |
| | 100 | 0.2802 | 1.1253 | 0.965 | 0.2853 | 1.1396 | 0.965 | 0.3228 | 1.2026 | 0.965 |
| | 200 | 0.1274 | 0.3725 | 0.954 | 0.1285 | 0.3741 | 0.954 | 0.1372 | 0.3785 | 0.954 |
| SPTE(0.50) | 40 | 0.5764 | 3.5665 | 0.961 | 0.6750 | 4.4558 | 0.961 | 0.8143 | 5.3086 | 0.961 |
| $\alpha = .50$ | 50 | 0.5499 | 2.8628 | 0.959 | 0.6008 | 3.2916 | 0.959 | 0.6673 | 3.5085 | 0.959 |
| | 100 | 0.2974 | 1.1785 | 0.965 | 0.3007 | 1.1899 | 0.965 | 0.3279 | 1.2283 | 0.965 |
| | 200 | 0.1320 | 0.3836 | 0.954 | 0.1320 | 0.3841 | 0.954 | 0.1469 | 0.3959 | 0.954 |
| SPTE(1.00) | 40 | 0.1180 | 1.9022 | 0.961 | 0.3009 | 3.2381 | 0.961 | 0.6845 | 4.9258 | 0.961 |
| $\alpha = .40$ | 50 | 0.2230 | 1.7830 | 0.959 | 0.3640 | 2.6332 | 0.959 | 0.5699 | 3.2558 | 0.959 |
| | 100 | 0.1781 | 0.9863 | 0.965 | 0.1902 | 1.0121 | 0.965 | 0.2709 | 1.1098 | 0.965 |
| | 200 | 0.0833 | 0.3189 | 0.954 | 0.0830 | 0.3222 | 0.954 | 0.1166 | 0.3399 | 0.954 |
| SPTE(1.00) | 40 | 0.1699 | 2.2238 | 0.961 | 0.3746 | 3.6742 | 0.961 | 0.7116 | 5.0504 | 0.961 |
| $\alpha = .45$ | 50 | 0.2877 | 2.1453 | 0.959 | 0.4171 | 2.8604 | 0.959 | 0.5884 | 3.3205 | 0.959 |
| | 100 | 0.2065 | 1.0532 | 0.965 | 0.2166 | 1.0723 | 0.965 | 0.2915 | 1.1563 | 0.965 |
| | 200 | 0.0984 | 0.3463 | 0.954 | 0.1007 | 0.3483 | 0.954 | 0.1180 | 0.3543 | 0.954 |
| SPTE(1.00) | 40 | 0.2634 | 2.8527 | 0.961 | 0.4605 | 4.0384 | 0.961 | 0.7392 | 5.1756 | 0.961 |
| $\alpha = .50$ | 50 | 0.3743 | 2.5568 | 0.959 | 0.4761 | 3.1285 | 0.959 | 0.6091 | 3.4178 | 0.959 |
| | 100 | 0.2408 | 1.1242 | 0.965 | 0.2473 | 1.1393 | 0.965 | 0.3017 | 1.1906 | 0.965 |
| | 200 | 0.1076 | 0.3611 | 0.954 | 0.1085 | 0.3617 | 0.954 | 0.1375 | 0.3775 | 0.954 |
| SE | 40 | 0.4197 | 2.3891 | 0.969 | 0.4925 | 2.9284 | 0.968 | 0.6708 | 3.8795 | 0.974 |
| | 50 | 0.3761 | 1.7427 | 0.969 | 0.4207 | 2.0360 | 0.966 | 0.5397 | 2.5275 | 0.973 |
| | 100 | 0.1845 | 0.7856 | 0.976 | 0.1932 | 0.8293 | 0.975 | 0.2442 | 0.9249 | 0.976 |
| | 200 | 0.0960 | 0.2662 | 0.974 | 0.0975 | 0.2701 | 0.973 | 0.1187 | 0.2851 | 0.973 |
| $SE^+$ | 40 | 0.4231 | 2.1692 | 0.977 | 0.4944 | 2.7118 | 0.976 | 0.6529 | 3.6658 | 0.981 |
| | 50 | 0.3888 | 1.6393 | 0.977 | 0.4324 | 1.9343 | 0.974 | 0.5368 | 2.4300 | 0.980 |
| | 100 | 0.2083 | 0.7112 | 0.983 | 0.2168 | 0.7553 | 0.982 | 0.2601 | 0.8556 | 0.983 |
| | 200 | 0.0959 | 0.2399 | 0.981 | 0.0973 | 0.2438 | 0.980 | 0.1153 | 0.2588 | 0.980 |

Table 2: NBIAS, SMSE and CP of the estimates for $k = 5$ with $\lambda = 0$ and $\beta^0 = 2$.

| | | WALD | SCT | LRT |
|---|---|---|---|---|
| | n | NBIAS SMSE CP | NBIAS SMSE CP | NBIAS SMSE CP |
| UMLE | 40 | 1.5202 11.6631 0.968 | 1.5202 11.6631 0.968 | 1.5202 11.6631 0.968 |
| | 50 | 1.3326 8.6147 0.969 | 1.3326 8.6147 0.969 | 1.3326 8.6147 0.969 |
| | 100 | 0.7496 3.5689 0.968 | 0.7496 3.5689 0.968 | 0.7496 3.5689 0.968 |
| | 200 | 0.3895 1.1946 0.966 | 0.3895 1.1946 0.966 | 0.3895 1.1946 0.966 |
| SRE(0.50) | 40 | 0.7601 2.9158 1.000 | 0.7601 2.9158 1.000 | 0.7601 2.9158 1.000 |
| | 50 | 0.6663 2.1537 1.000 | 0.6663 2.1537 1.000 | 0.6663 2.1537 1.000 |
| | 100 | 0.3748 0.8922 1.000 | 0.3748 0.8922 1.000 | 0.3748 0.8922 1.000 |
| | 200 | 0.1948 0.2987 1.000 | 0.1948 0.2987 1.000 | 0.1948 0.2987 1.000 |
| SPTE(0.50) | 40 | 0.7437 3.3019 0.968 | 0.7761 3.8449 0.968 | 0.8578 4.6054 0.974 |
| $\alpha = .10$ | 50 | 0.6528 2.4034 0.969 | 0.6778 2.7931 0.969 | 0.7655 3.4871 0.976 |
| | 100 | 0.3640 1.0438 0.968 | 0.3824 1.2370 0.968 | 0.4875 1.9620 0.970 |
| | 200 | 0.1981 0.3961 0.966 | 0.2025 0.4304 0.966 | 0.2562 0.6032 0.968 |
| SPTE(0.50) | 40 | 0.7477 3.4201 0.968 | 0.8041 4.2110 0.968 | 0.9167 5.3375 0.969 |
| $\alpha = .15$ | 50 | 0.6616 2.6412 0.966 | 0.6960 3.0396 0.969 | 0.8091 3.9858 0.974 |
| | 100 | 0.3708 1.1216 0.968 | 0.4025 1.4525 0.968 | 0.5345 2.3175 0.968 |
| | 200 | 0.2057 0.4628 0.969 | 0.2143 0.5023 0.966 | 0.2824 0.7185 0.966 |
| SPTE(0.50) | 40 | 0.7507 3.5156 0.968 | 0.8101 4.3670 0.968 | 0.9712 6.1134 0.969 |
| $\alpha = .20$ | 50 | 0.6721 2.8245 0.969 | 0.7108 3.2559 0.969 | 0.9046 5.1269 0.969 |
| | 100 | 0.3775 1.2204 0.968 | 0.4188 1.5909 0.968 | 0.5714 2.5466 0.968 |
| | 200 | 0.2208 0.5270 0.966 | 0.2330 0.5736 0.966 | 0.3004 0.7966 0.966 |
| SPTE(1.00) | 40 | 0.0395 0.5148 0.968 | 0.0361 1.2388 0.968 | 0.1967 2.2528 0.974 |
| $\alpha = .10$ | 50 | 0.0311 0.3329 0.969 | 0.0293 0.8525 0.969 | 0.1987 3.4871 0.976 |
| | 100 | 0.0238 0.2021 0.968 | 0.0189 0.4597 0.968 | 0.2254 1.7779 0.970 |
| | 200 | 0.0091 0.1299 0.966 | 0.0172 0.1756 0.966 | 0.1229 0.4061 0.968 |
| SPTE(1.00) | 40 | 0.0315 0.6724 0.968 | 0.0889 1.7269 0.968 | 0.3137 3.2290 0.969 |
| $\alpha = .15$ | 50 | 0.0240 0.6500 0.966 | 0.0616 1.1813 0.969 | 0.2862 2.4429 0.974 |
| | 100 | 0.0133 0.3058 0.968 | 0.0562 0.7470 0.968 | 0.3195 1.9004 0.968 |
| | 200 | 0.0236 0.2188 0.969 | 0.0394 0.2715 0.966 | 0.1754 0.5597 0.966 |
| SPTE(1.00) | 40 | 0.0303 0.7998 0.968 | 0.1010 1.9349 0.968 | 0.4225 4.2635 0.969 |
| $\alpha = .20$ | 50 | 0.0272 0.8944 0.969 | 0.0924 1.4697 0.969 | 0.4768 3.9643 0.969 |
| | 100 | 0.0169 0.4376 0.968 | 0.0894 0.9316 0.968 | 0.3933 2.2058 0.968 |
| | 200 | 0.0524 0.3044 0.966 | 0.0766 0.3665 0.966 | 0.2113 0.6640 0.966 |
| SE | 40 | 0.1406 2.5705 0.981 | 0.1542 2.9850 0.981 | 0.6417 4.0332 0.984 |
| | 50 | 0.0933 1.9388 0.985 | 0.1411 2.2813 0.986 | 0.5584 3.0446 0.989 |
| | 100 | 0.1274 0.8349 0.988 | 0.1921 1.0696 0.986 | 0.3725 1.5256 0.987 |
| | 200 | 0.1017 0.3762 0.988 | 0.0533 0.4033 0.985 | 0.1892 0.4854 0.986 |
| $SE^+$ | 40 | 0.2036 1.1969 0.987 | 0.3762 2.0855 0.987 | 0.6988 3.5936 0.989 |
| | 50 | 0.1985 0.9392 0.996 | 0.3399 1.5925 0.995 | 0.6121 2.6847 0.999 |
| | 100 | 0.2004 0.6328 0.997 | 0.2567 0.8805 0.995 | 0.3994 1.3705 0.997 |
| | 200 | 0.1407 0.2855 0.998 | 0.1515 0.3147 0.996 | 0.2080 0.4080 0.996 |

Table 3: NBIAS, SMSE and CP of the estimates for $k = 5$ with $\beta^0 = 1$.

| | | $\lambda = 0$ | | | $\lambda = -7$ | | |
|---|---|---|---|---|---|---|---|
| | n | NBIAS | SMSE | CP | NBIAS | SMSE | CP |
| UMLE | 40 | 0.7951 | 5.0308 | 0.978 | 1.2970 | 17.6507 | 0.979 |
| | 50 | 0.5743 | 2.7477 | 0.975 | 1.0810 | 12.6502 | 0.959 |
| | 100 | 0.2539 | 0.7545 | 0.975 | 0.4672 | 3.6295 | 0.957 |
| | 200 | 0.1039 | 0.3009 | 0.962 | 0.1812 | 1.2642 | 0.957 |
| SRE(0.50) | 40 | 0.3976 | 1.2577 | 1.000 | 0.6485 | 4.4127 | 1.000 |
| | 50 | 0.2872 | 0.6869 | 1.000 | 0.5405 | 3.1625 | 1.000 |
| | 100 | 0.1269 | 0.1886 | 1.000 | 0.2336 | 0.9074 | 1.000 |
| | 200 | 0.0520 | 0.0752 | 1.000 | 0.0906 | 0.3161 | 1.000 |
| SPTE(0.50) | 40 | 0.5594 | 3.3585 | 0.978 | 0.8567 | 9.4953 | 0.984 |
| $\alpha = .10$ | 50 | 0.3892 | 1.6519 | 0.976 | 0.6725 | 6.9649 | 0.963 |
| | 100 | 0.1668 | 0.3807 | 0.976 | 0.3112 | 1.8927 | 0.961 |
| | 200 | 0.0657 | 0.1398 | 0.962 | 0.1114 | 0.5739 | 0.958 |
| SPTE(0.50) | 40 | 0.6128 | 3.7792 | 0.978 | 0.8887 | 10.8005 | 0.980 |
| $\alpha = .15$ | 50 | 0.4126 | 1.8163 | 0.975 | 0.7382 | 8.0480 | 0.960 |
| | 100 | 0.1757 | 0.4249 | 0.975 | 0.3360 | 2.1821 | 0.957 |
| | 200 | 0.0679 | 0.1591 | 0.962 | 0.1188 | 0.6534 | 0.957 |
| SPTE(0.50) | 40 | 0.6381 | 4.0036 | 0.978 | 0.9468 | 11.9207 | 0.970 |
| $\alpha = .20$ | 50 | 0.4394 | 1.9812 | 0.975 | 0.7982 | 9.0435 | 0.959 |
| | 100 | 0.1793 | 0.4525 | 0.975 | 0.3512 | 2.3573 | 0.957 |
| | 200 | 0.0722 | 0.1765 | 0.962 | 0.1264 | 0.7266 | 0.957 |
| SPTE(1.00) | 40 | 0.3246 | 2.8010 | 0.978 | 0.4186 | 6.7769 | 0.984 |
| $\alpha = .10$ | 50 | 0.2049 | 1.2867 | 0.976 | 0.2676 | 5.0698 | 0.963 |
| | 100 | 0.0799 | 0.2561 | 0.976 | 0.1566 | 1.3138 | 0.961 |
| | 200 | 0.0279 | 0.0860 | 0.962 | 0.0425 | 0.3438 | 0.958 |
| SPTE(1.00) | 40 | 0.4310 | 3.3620 | 0.978 | 0.4829 | 8.5171 | 0.980 |
| $\alpha = .15$ | 50 | 0.2517 | 1.5058 | 0.975 | 0.3995 | 6.5140 | 0.960 |
| | 100 | 0.0980 | 0.3151 | 0.975 | 0.2060 | 1.6996 | 0.957 |
| | 200 | 0.0322 | 0.1119 | 0.962 | 0.0571 | 0.4497 | 0.957 |
| SPTE(1.00) | 40 | 0.4815 | 3.6612 | 0.978 | 0.5991 | 10.0107 | 0.979 |
| $\alpha = .20$ | 50 | 0.3051 | 1.7257 | 0.975 | 0.5163 | 7.8413 | 0.959 |
| | 100 | 0.1052 | 0.3518 | 0.975 | 0.2358 | 1.9333 | 0.957 |
| | 200 | 0.0412 | 0.1350 | 0.962 | 0.0723 | 0.5473 | 0.957 |
| SE | 40 | 0.4424 | 2.4204 | 0.989 | 0.6449 | 7.2347 | 0.991 |
| | 50 | 0.3071 | 1.2188 | 0.989 | 0.5158 | 5.3482 | 0.990 |
| | 100 | 0.1070 | 0.3066 | 0.987 | 0.2330 | 1.6733 | 0.984 |
| | 200 | 0.0485 | 0.1228 | 0.983 | 0.0738 | 0.5133 | 0.984 |
| $SE^+$ | 40 | 0.4612 | 2.3030 | 0.997 | 0.6979 | 6.7316 | 0.998 |
| | 50 | 0.3175 | 1.1360 | 0.994 | 0.5550 | 4.9396 | 0.996 |
| | 100 | 0.1235 | 0.2494 | 0.998 | 0.2578 | 1.3288 | 0.995 |
| | 200 | 0.0510 | 0.0941 | 0.994 | 0.0891 | 0.3826 | 0.995 |

is used. For the choice of different test statistics, we observed differences in the CPs of the estimators SE and SE$^+$ with $k = 3$ (5) that does not follow any regular pattern though.

The results discussed above hold for the case of $\beta^0 = 2$. Comparing this case (for $k = 5$ with the LRT) with $\beta^0 = 2$, we observed that the former case produces larger CPs than the latter for all the estimators.

Table 4: Sampling properties of the test statistics with $\lambda = 0$ and $\beta^0 = 2$.

| $k=3$ | | | | | | | |
|---|---|---|---|---|---|---|---|
| | | | | $\alpha = .05$ | $\alpha = .40$ | $\alpha = .45$ | $\alpha = .50$ |
| | n | Mean sd | Mean se<br>$p$-value | PSIG se | PSIG se | PSIG se | PSIG se |
| WALD | 40 | 2.66 2.408 | .54799 .016038 | .0395 .00627 | .2970 .01471 | .3396 .01528 | .3905 .01571 |
| | 50 | 2.76 2.364 | .52585 .015902 | .0406 .00628 | .3448 .01513 | .4047 .01563 | .4574 .01587 |
| | 100 | 2.95 2.327 | .49791 .015811 | .0350 .00581 | .4050 .01552 | .4550 .01575 | .5110 .01581 |
| | 200 | 2.93 2.356 | .50256 .015811 | .0460 .00662 | .3900 .01542 | .4450 .01572 | .4870 .01581 |
| SCT | 40 | 3.06 2.997 | .51632 .016104 | .0665 .00803 | .3458 .01533 | .3967 .01576 | .4538 .01604 |
| | 50 | 3.07 2.798 | .49901 .015923 | .0568 .00737 | .3945 .01556 | .4452 .01583 | .4959 .01592 |
| | 100 | 3.11 2.554 | .48626 .015805 | .0510 .00696 | .4180 .01560 | .4700 .01578 | .5250 .01579 |
| | 200 | 3.00 2.467 | .49757 .015811 | .0480 .00676 | .4000 .01549 | .4490 .01573 | .4910 .01581 |
| LRT | 40 | 3.18 2.667 | .48438 .016104 | .0602 .00767 | .4299 .01595 | .4683 .01608 | .5130 .01611 |
| | 50 | 3.22 2.506 | .47137 .015897 | .0629 .00779 | .4442 .01582 | .4757 .01590 | .5325 .01589 |
| | 100 | 3.24 2.600 | .47177 .015786 | .0580 .00739 | .4460 .01572 | .5000 .01581 | .5410 .01576 |
| | 200 | 3.08 2.505 | .48997 .015808 | .0500 .00689 | .4080 .01554 | .4530 .01574 | .5180 .01581 |
| $k=5$ | | | | | | | |
| | | | | $\alpha = .05$ | $\alpha = .10$ | $\alpha = .15$ | $\alpha = .20$ |
| WALD | 40 | 4.13 3.189 | .59894 .018005 | .0324 .00650 | .0594 .00868 | .0837 .01017 | .1039 .01121 |
| | 50 | 4.15 2.751 | .58584 .016925 | .0307 .00593 | .0626 .00832 | .0921 .00994 | .1240 .01132 |
| | 100 | 4.59 2.862 | .53314 .015856 | .0323 .00562 | .0687 .00804 | .0929 .00923 | .1404 .01104 |
| | 200 | 4.79 2.870 | .51077 .015808 | .0340 .00573 | .0760 .00838 | .1230 .01039 | .1690 .01185 |
| SCT | 40 | 5.07 4.465 | .53113 .018332 | .0675 .00922 | .1066 .01134 | .1431 .01286 | .1741 .01393 |
| | 50 | 4.89 3.531 | .52660 .017156 | .0685 .00868 | .1063 .01059 | .1429 .01202 | .1842 .01332 |
| | 100 | 5.02 3.298 | .50119 .015891 | .0556 .00728 | .0889 .00904 | .1404 .01104 | .1859 .01236 |
| | 200 | 5.00 3.089 | .49642 .015811 | .0490 .00683 | .0960 .00927 | .1430 .01107 | .1940 .01250 |
| LRT | 40 | 5.05 3.131 | .49199 .018366 | .0513 .00810 | .0891 .01046 | .1417 .01281 | .1916 .01446 |
| | 50 | 5.06 3.018 | .48698 .017174 | .0460 .00720 | .0909 .00988 | .1358 .01173 | .2161 .01414 |
| | 100 | 5.30 3.274 | .47180 .015866 | .0546 .00722 | .1162 .01014 | .1838 .01231 | .2323 .01340 |
| | 200 | 5.23 3.220 | .47821 .015796 | .0550 .00721 | .1210 .01031 | .1840 .01225 | .2380 .01347 |

So far we have discussed above the results about the case of $\lambda = 0$. Considering a single case where $\lambda = -7$ (for $k = 5$ with the LRT) with $\beta^0 = 1$, we found that the CPs are larger

Table 5: Sample quantiles of the test statistics with $\lambda = 0$ and $\beta^0 = 2$.

| $k = 3$ | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | n | 10% | 20% | 30% | 40% | 50% | 60% | 70% | 80% | 90% | 95% | 99% |
| WALD | 40 | 0.59 | 0.97 | 1.30 | 1.65 | 1.99 | 2.34 | 2.89 | 3.86 | 5.83 | 7.36 | 11.34 |
| | 50 | 0.62 | 1.04 | 1.45 | 1.80 | 2.13 | 2.67 | 3.19 | 3.92 | 5.30 | 7.14 | 12.84 |
| | 100 | 0.62 | 1.06 | 1.54 | 1.95 | 2.40 | 2.98 | 3.59 | 4.59 | 5.98 | 7.25 | 11.36 |
| | 200 | 0.63 | 1.04 | 1.47 | 1.90 | 2.31 | 2.91 | 3.53 | 4.46 | 5.84 | 7.48 | 11.07 |
| SCT | 40 | 0.60 | 1.00 | 1.38 | 1.74 | 2.20 | 2.62 | 3.32 | 4.50 | 6.80 | 8.71 | 14.71 |
| | 50 | 0.63 | 1.07 | 1.50 | 1.89 | 2.33 | 2.91 | 3.61 | 4.42 | 6.00 | 8.18 | 15.01 |
| | 100 | 0.57 | 1.08 | 1.57 | 1.98 | 2.48 | 3.08 | 3.77 | 4.88 | 6.48 | 7.84 | 12.45 |
| | 200 | 0.64 | 1.05 | 1.48 | 1.92 | 2.33 | 2.95 | 3.60 | 4.59 | 6.08 | 7.64 | 11.40 |
| LRT | 40 | 0.61 | 1.46 | 1.38 | 1.92 | 2.43 | 3.14 | 3.82 | 5.11 | 6.56 | 8.29 | 12.64 |
| | 50 | 0.66 | 1.59 | 1.50 | 2.03 | 2.54 | 3.30 | 4.10 | 5.00 | 6.51 | 8.39 | 11.37 |
| | 100 | 0.57 | 1.57 | 1.57 | 2.12 | 2.64 | 3.29 | 4.04 | 5.03 | 6.43 | 7.24 | 12.49 |
| | 200 | 0.62 | 1.49 | 1.48 | 1.93 | 2.43 | 3.00 | 3.68 | 4.73 | 6.24 | 7.84 | 13.10 |
| TQ* | | 0.58 | 1.01 | 1.42 | 1.87 | 2.37 | 2.95 | 3.66 | 4.64 | 6.25 | 7.81 | 11.34 |
| $k = 5$ | | | | | | | | | | | | |
| WALD | 40 | 1.43 | 2.01 | 2.41 | 2.91 | 3.34 | 3.84 | 4.57 | 5.78 | 7.34 | 9.82 | 19.32 |
| | 50 | 1.51 | 2.01 | 2.51 | 2.98 | 3.48 | 4.11 | 4.75 | 5.80 | 7.90 | 9.90 | 14.22 |
| | 100 | 1.55 | 2.25 | 2.92 | 3.55 | 4.13 | 4.70 | 5.39 | 6.44 | 7.97 | 10.29 | 14.41 |
| | 200 | 1.58 | 2.35 | 3.08 | 3.67 | 4.26 | 4.97 | 5.86 | 6.93 | 8.68 | 10.21 | 14.04 |
| SCT | 40 | 1.55 | 2.18 | 2.75 | 3.28 | 3.96 | 4.56 | 5.55 | 6.81 | 9.54 | 12.21 | 25.34 |
| | 50 | 1.63 | 2.18 | 2.74 | 3.32 | 4.04 | 4.75 | 5.52 | 7.07 | 9.40 | 12.04 | 17.59 |
| | 100 | 1.60 | 2.34 | 3.07 | 3.72 | 4.43 | 5.14 | 5.92 | 7.02 | 8.89 | 11.36 | 16.89 |
| | 200 | 1.58 | 2.38 | 3.15 | 3.76 | 4.39 | 5.17 | 6.10 | 7.22 | 9.10 | 11.05 | 14.95 |
| LRT | 40 | 1.76 | 2.40 | 3.11 | 3.75 | 4.44 | 5.33 | 6.15 | 7.12 | 9.05 | 11.23 | 15.43 |
| | 50 | 1.71 | 2.52 | 3.11 | 3.79 | 4.53 | 5.26 | 6.18 | 7.40 | 8.93 | 10.91 | 14.70 |
| | 100 | 1.67 | 2.45 | 3.16 | 3.99 | 4.79 | 5.52 | 6.48 | 7.72 | 9.63 | 11.31 | 16.06 |
| | 200 | 1.66 | 2.41 | 3.14 | 3.84 | 4.60 | 5.41 | 6.49 | 7.78 | 9.74 | 11.28 | 14.66 |
| TQ* | | 1.61 | 2.34 | 3.00 | 3.66 | 4.35 | 5.13 | 6.06 | 7.29 | 9.24 | 11.07 | 15.09 |

* = Theoretical Quantiles

than the case of its counterpart $\lambda = 0$. Ordering of the estimators according to their CPs looks like $\text{SE}^+ > \text{SE} > \text{SPTE} \geq \text{UMLE}$.

## 3.5 Test Statistics

The mean values of the test statistics strictly follow the inequality WALD $<$ SCT $<$ LRT with $k = 3$ (5), with SCT for the sample size 40 ($k = 5$) the only exception in the present case. Considering individual samples we find that in almost 60% of the samples the values of the test statistics follow the inequality. Roughly speaking, in almost all samples the WALD statistic gives a smaller value than the SCT, in an around 60-70% samples the WALD is smaller than the LRT while in an around 55-60% samples the SCT possesses a smaller value than the LRT. The distributions of the test statistics (with degrees of freedom 3 (5)) clearly represent a chi-square distribution with the corresponding degrees of freedom (see Figure 3, Figure 4).

For the proportion significant (PSIG) of the test statistics, the inequality clearly holds at all level of significance considered with $k = 3$ (5). No sharp trend in the values of the PSIG is found for an increase in the sample size. However, in large samples the PSIG is close enough to the value what it would be for the corresponding level of significance considered. In this criterion, the SCT and LRT behaves more or less in a similar fashion and give more closer result to (.05, .40, .45, .50) ((.05, .10, .15, .20)) with $k = 3$ (5). On some occasions (with higher significance level) the SCT is found to be well-performed compared to the LRT. However, the performance of the WALD statistic is not that good.

With $k = 5$, the quantiles of the test statistics give the inequality WALD $<$ SCT $<$ LRT with few exceptions in the higher quantiles. However, the inequality holds with few exceptions in the lower and higher quantiles for the case with $k = 3$. No recognizable difference is observed for an increase in the sample size. However, in large samples the test statistics behaves more or less in a similar fashion. The SCT and LRT give closer results. The performance of the WALD statistic can easily be differentiated from the other two. As expected, the empirical quantiles of the test statistics tend to approximate the theoretical quantiles very well with few exceptions in the larger quantiles.

It is interesting to note that the behavior of the test statistics that we observed here is totally in agreement with the results in Matin (2005) though in different context.

### 3.6 Graph Analysis

**Frequency Histogram for Estimators:** Frequency histograms for UMLE, SRE($c = .50$), SPTE ($c = .50$), SPTE($c = 1.00$), SE and SE$^+$ are displayed in Figure 1 and Figure 2 for the sample sizes 40, 50, 100, and 200 for only $\beta_1$ ($\beta_3$) with $k = 3$ (5) with $\lambda = 0$ where the LRT is used. Examination of the histograms reveals that the distribution of the unrestricted estimator is negatively skewed, but approaches a normal curve as the sample size increases. However, for the other estimators the distribution resembles a symmetric curve with a very steep peakedness, and a slightly longer tail on the right. This nature of the curves become smoother as the sample size increases, but in no way approaches a normal curve. (So, it may be concluded that as the sample size increases the estimators approach toward the true value.)

**Frequency Histogram for Test Statistics:** Frequency histograms for the WALD, SCT and LRT are displayed in Figure 3 and Figure 4 for the sample sizes 40, 50, 100, and 200 for $k = 3$ (5). It is clear that as the sample size increases the distribution of the test statistics approximates the chi-square distribution nicely for both $k = 3$ and 5. Close examination of the histograms reveals the differences among test statistics in small

Figure 1: Frequency histogram of the regression coefficient $\beta_1$ for the estimators UMLE, SRE($c = .50$), SPTE($c = .50$), SPTE($c = 1.00$), SE and SE$^+$ with $k = 3$ while the LRT is used to test $H_0 : \beta_0 = \beta_1 = \beta_2 = 2$.

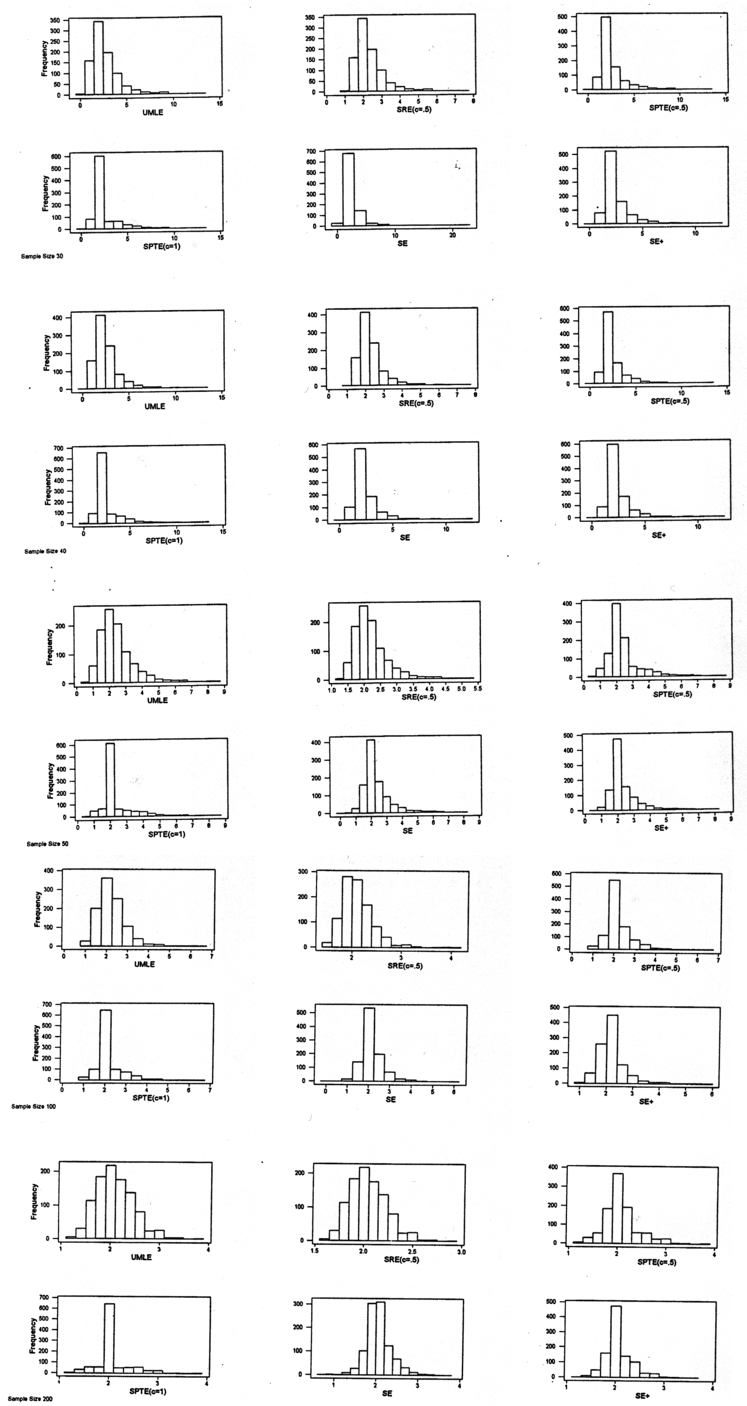
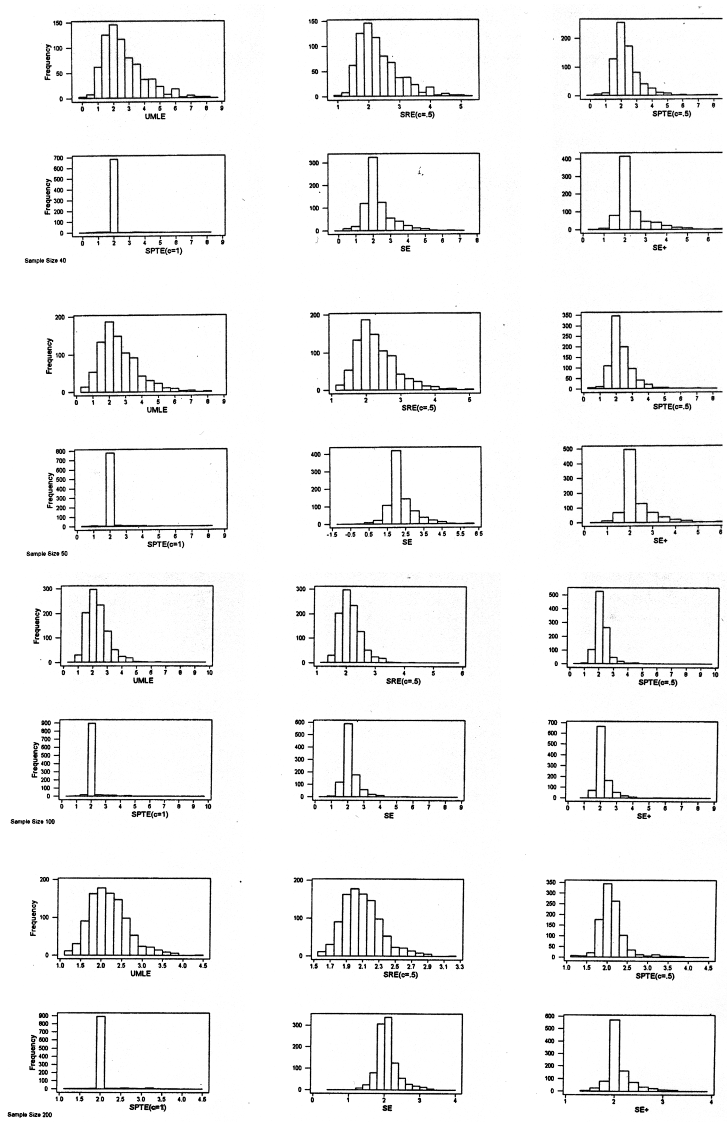


Figure 2: Frequency histogram of the regression coefficient $\beta_3$ for the estimators UMLE, SRE($c = .50$), SPTE($c = .50$), SPTE($c = 1.00$), SE and SE$^+$ with $k = 5$ while the LRT is used to test $H_0 : \beta_0 = \beta_1 = \beta_2 = \beta_3 = \beta_4 = 2$.



samples. The impact of an increase in the number of explanatory variables can be seen in the histogram of the sample size 200. If we compare the histograms of the sample size 200 for $k = 3$ with the same for $k = 5$ we see that the convergence of the distribution to chi-square is much smoother in $k = 3$ than in $k = 5$. This indicates that a much larger

Figure 3: Frequency histogram of the test statistics WALD, SCT and LRT while used to test $H_0 : \beta_0 = \beta_1 = \beta_2 = 2$ with $k = 3$.

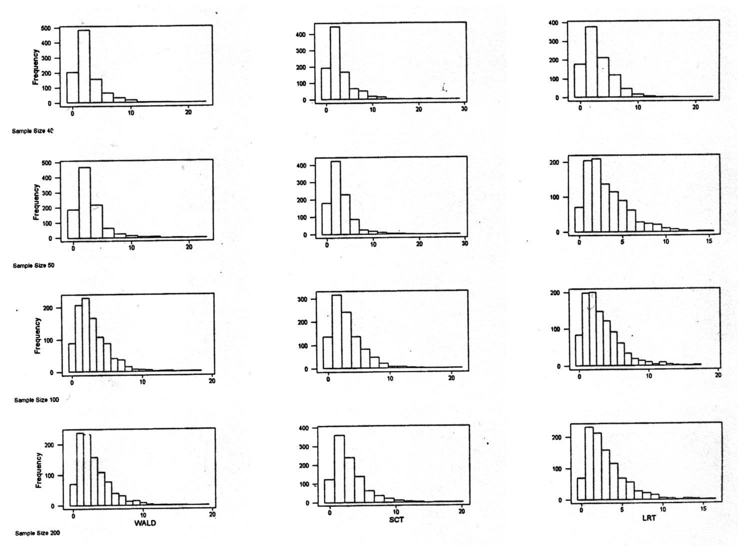
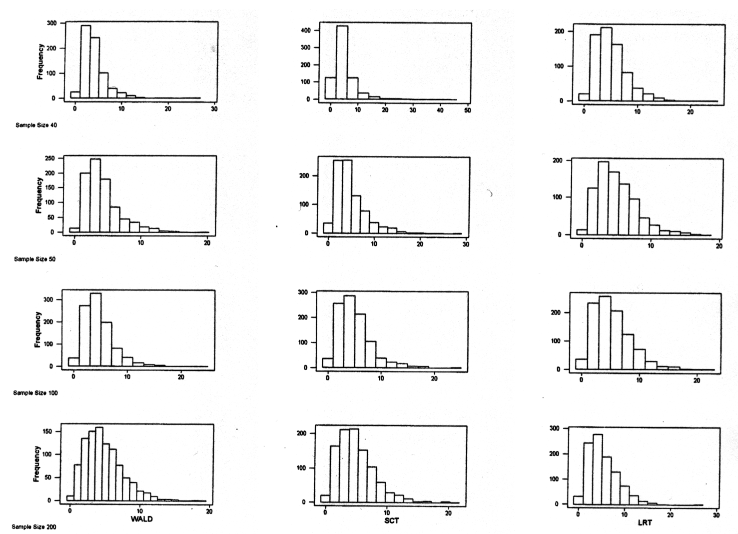


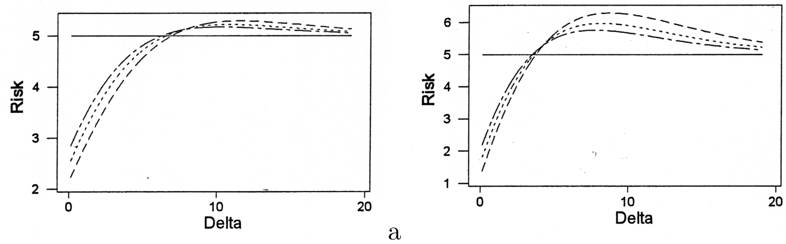

Figure 4: Frequency histogram of the test statistics WALD, SCT and LRT while used to test $H_0 : \beta_0 = \beta_1 = \beta_2 = \beta_3 = \beta_4 = 2$ with $k = 5$.




sample size is necessary to have a close approximation to the chi-square distribution for the case of $k = 5$.

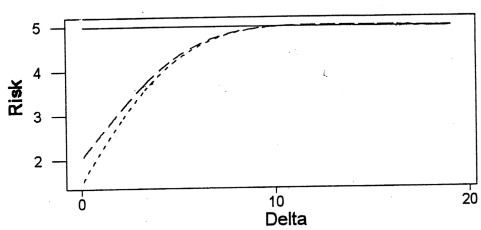
**Risk Graph for Estimators:** From the formulas (4.27-4.31 in Matin and Saleh (2005))

Figure 5: Risk curves for the estimators UMLE, SPTE($c = .50$) and SPTE ($c = 1.00$) for $k = 5$. The solid line (__) represents the UMLE. The dashed line (- - - -), dotted line (. . . .) and dash 1-dotted line (- . - . - .) represent the significance level $\alpha = .10$, $\alpha = .15$ and $\alpha = .20$ respectively, for SPTE($c = .50$) in (a) and for SPTE($c = 1.00$) in (b).




a b

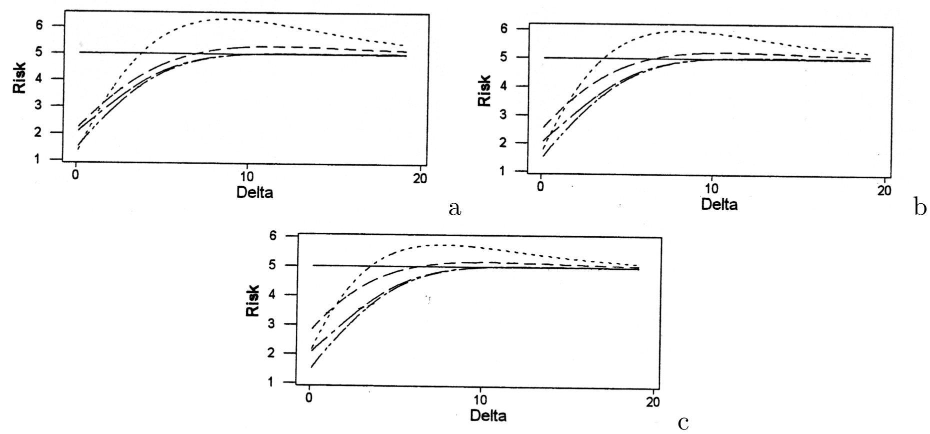
Figure 6: Risk curves for the estimators UMLE, SE and SE$^+$ for $k = 5$. The solid line (__), dashed line (- - - -) and dotted line (. . . .) represent the UMLE, SE and SE$^+$ respectively.




of the risks of the estimators it is clear that their values depend on the two matrices $Q$ and $D_1^{-1}$. For an ideal situation, we let $Q = D_1^{-1}$ so that $tr(QD_1) = tr(\mathbf{I}_k) = k$. Thus the risk of the unrestricted estimator becomes $k$ and the other formulas in 4.28-4.31 changed accordingly and are given in 4.32-4.36 in Matin and Saleh(2005). These risks for the estimators were computed and used in for drawing risk graphs as a function of the noncentrality parameter, $\Delta$. And comparisons are made with the risk line of $k = 5$ for the unrestricted estimator UMLE.

In Figure 5(a, b), we present the risk graphs for SPTE(c = .50) in (a) and SPTE(c = 1.00) in (b) to see how the levels of significance effect the risk level of the estimators. It is observed that at all significance levels the risk curves in (a) ((b)) cross the line ($k = 5$) and intersect each other at a certain level of $\Delta$ and spread out again but never cross the line. The co-ordinates of this intersection point are not the same in (a) and (b). We could see that the distance between the $y$-axis and the intersection point in (a) is larger than in (b). A closer look tells that the three curves corresponding to the three levels of significance, do not show the same behavior before and after the intersection point in (a) and (b). The curve for the optimal level of significance remains in the middle however, the other two interchanges their position after the intersection point. It is also clear that the curves are very much closer to the line ($k = 5$) in (a). The superiority of SPTE over the unrestricted estimator up to a certain level of $\Delta$ is quite clear from the graph while the graph for SPTE(c = .50)

Figure 7: Risk curves for the estimators UMLE, SPTE($c = .50$), SPTE ($c = 1.00$), SE and SE$^+$ for $k = 5$. The solid line (__), dashed line (- - - -), dotted line (. . . .), dash 1-dotted line ( -.-.-.-.- ) and dash 2-dotted line (- .. - .. - ..) represent the UMLE, SPTE($c = .50$), SPTE($c = 1.00$), SE and SE$^+$ respectively, in (a) for the significance level $\alpha = .10$, in (b) for the significance level $\alpha = .15$ and in (c) for the significance level $\alpha = .20$.




starts at a higher risk level than SPTE(c = 1.00).

Figure 8: Risk curves for the estimators UMLE, SRE($c = .50$), SPTE($c = .50$), SPTE($c = 1.00$), SE and SE+ for $k = 5$. The solid line (__), dashed line (- - - -), dotted line (. . . .), dash 1-dotted line (- . - - . - .), dash 2-dotted line (- .. - .. - ..) and dash 3-dotted line (- ... - ... - ...) represent the UMLE, SRE($c = .50$), SPTE($c = .50$), SPTE($c = 1.00$), SE and SE$^+$ respectively, in (a) for the significance level $\alpha = .10$, in (b) for the significance level $\alpha = .15$ and in (c) for the significance level $\alpha = .20$.

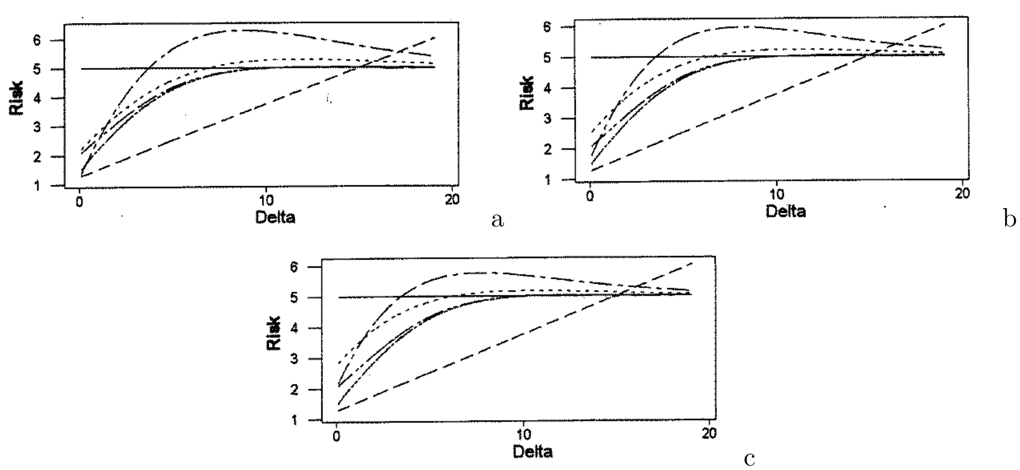


Figure 6 displays the risk curves for SE and SE$^+$ along with the risk line ($k = 5$). The superiority of SE and SE$^+$ over the unrestricted estimator is clearly observed. Furthermore, the dominance of SE$^+$ over SE is also seen from the graph. It is observed that SE (SE$^+$) begins at 2 (0) and approaches to the risk line at $k = 5$ while the gap between them decreases as $\Delta$ increases. At a certain level of $\Delta$ they nearly touch the line and both maintain the same space and run along the line.

Figure 7 (a, b, c) displays the risk curves for SPTE(c = .50), SPTE(c = 1.00), SE and SE$^+$ along with the risk line of the unrestricted estimator for $\alpha = .10$ in (a), for $\alpha = .15$ in (b), and for $\alpha = .20$ in (c). These graphs allow us to see the superiority picture of the estimators along the line of change in the level of significance in SPTE. Here, one can easily identify the superiority of SPTE(c = .50) over the SPTE(c = 1.00) at all the three levels of significance. Other comments in the analysis of Figure 5 and Figure 6 remain valid for this case also.

Figure 9: Risk curves for the estimators UMLE, SRE($c$ = .50), SRE($c$ = 1.00), SPTE ($c$ = .50), SPTE($c$ = 1.00), SE and SE$^+$ for $k = 5$. The solid line (__) UMLE, dashed line (- - - -), dotted line (. . . . .), dash 1-dotted line (- . - . - .), dash 2-dotted line (- .. - .. - ..), dash 3-dotted line (- ... - ... - ...) and long dashed line (— — —) represent the UMLE, SRE($c$ = .50), SRE($c$ = 1.00), SPTE($c$ = .50), SPTE($c$ = 1.00), SE and SE$^+$ respectively in (a) for the significance level $\alpha = .10$, in (b) for the significance level $\alpha = .15$ and in (c) for the significance level $\alpha = .20$.

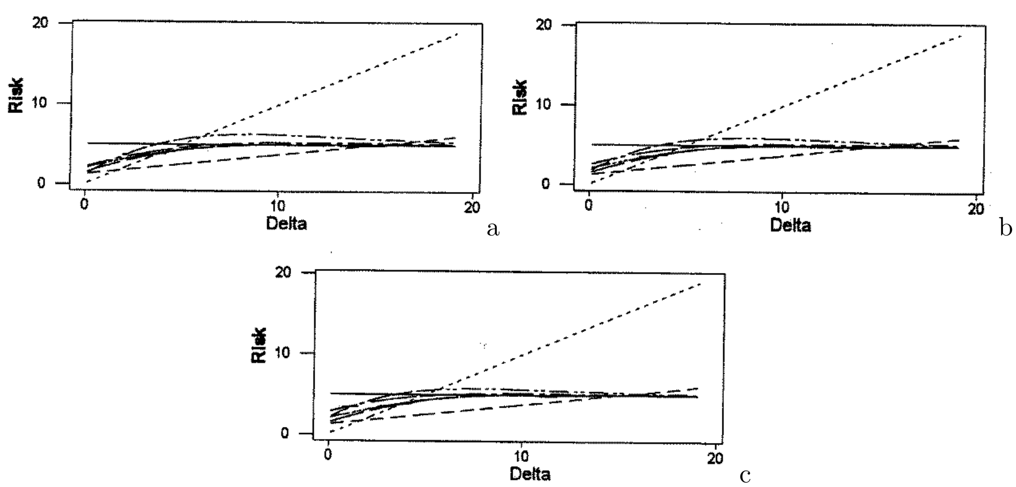



Figure 8 is nothing but an extension of Figure 7 with the inclusion of the risk curve for SRE(c = .50). For $\Delta = 0$ (i.e., under null hypothesis) the risk of SRE(c = .50) is the smallest one compared to others. However, as $\Delta$ increases it also increases unboundedly crossing the risk line at a certain level of $\Delta$. Figure 9 is a further extension of Figure 8 with the inclusion of the risk curve SRE(c = 1.00). It gives the smallest risk at $\Delta = 0$ however, its tremendous unboundedness makes the difference between the other curves smaller.

# 4 Summary

It is theoretically possible to have larger bias for the estimators compared to the UMLE however, no such case is encountered. In general, the estimators reduce the bias inherent in the UMLE by approximately 10-50%. Note that, an increase in the number of explanatory variables shows an increase in the bias of the estimators. For SPTE (where the level of significance is directly involved), for an increase within the chosen levels of significance the bias also increases. For SE and $SE^+$ (where the value of the test statistic is directly involved), we observe a regular pattern $\text{Bias}_{LRT} > \text{Bias}_{SCT} > \text{Bias}_{WALD}$. This is also true for SPTE. As the size of the parameter value (i.e., the true value of the regression coefficient) tend towards zero the bias decreases. An increase in the degree of skewness in the explanatory variables leads to an increase in the bias of the estimators. In large samples, the bias for all of the estimators is remarkably close to zero.

In general, as the sample size increases the mean square error for all of the estimators decreases. Note that, the estimators have smaller MSE compared to the UMLE. An increase in the number of explanatory variables leads to an increase in the MSE, too. For SPTE (where the level of significance is directly involved), for a decrease within the chosen levels of significance the MSE also decreases. For SE and $SE^+$ (where the value of the test statistic is directly involved), we observe a regular pattern $\text{MSE}_{LRT} > \text{MSE}_{SCT} > \text{MSE}_{WALD}$. This is also true for SPTE. As the size of the parameter value (i.e., the true value of the regression coefficient) increases the MSE increases too. An increase in the degree of skewness of the explanatory variables leads to an increase in the MSE too. In large samples, the MSE for all of the estimators is remarkably close to each other.

In general, as the sample size increases the coverage probabilities for the estimators UMLE and SPTE tend toward the nominal level. Note that, the SE and $SE^+$ possess larger CPs as expected compared to UMLE and SPTE. An increase in the number of explanatory variables leads to an increase in CPs also. For SPTE (where the level of significance is directly involved), we do not observe any significant effect on the CPs of the estimators for an increase or decrease within the chosen level of significance. For SE and $SE^+$ (where the value of the test statistic is directly involved), we observe differences in the CPs of the estimators however, that does not follow any regular pattern. As the size of the parameter value (i.e., the true value of the regression coefficient) increases the CPs decrease slightly. Increase in the degree of skewness of the explanatory variables shows larger CPs compared to its counterpart.

Considering the sampling properties of the test statistics the LRT stands out as the best. However, the strict inequality WALD $<$ SCT $<$ LRT holds for the mean value of the test statistics. This confirms Matin (2005) results. In large samples, all the three test statistics give identical results and quite nicely approximate the chi-square distribution with appropriate degrees of freedom.

Our simulation study thus confirms the theoretical results of Matin and Saleh (2005) in cases where comparisons are possible. It is clear that when the number of explanatory variables is greater than or equal to 3 the shrinkage and the positive-rule shrinkage estimators *always* perform well. Considering the MSE the positive-rule shrinkage estimator performs better than the shrinkage estimator. The sampling properties of the test statistics indicate that the likelihood ratio test is the best. However, we lean toward the use of the Wald statistic when the problem of estimation is of paramount interest as it provides lower bias and lower MSE for the estimators compared to the score statistic and likelihood ratio statistic.

## Acknowledgement

The work of both the authors sponsored by the NSERC Grant No. A3088, Canada. The first author was also partially funded by the Department of Statistics, Uppsala University, Sweden. Authors would like to thank Reinhold Bergström for his valuable comments on an earlier version of the paper.

## References

[1] Azzalini, A. (1985). A class of distributions which includes the normal ones, *Scandinavian Journal of Statistics,* **12***, 171-178.*

[2] George, E. I. and Casella, G. (1994). An empirical Bayes confidence report, *Statistica Sinica,* **4***, 617-638.*

[3] Griffiths, P. and Hill, I. D. (1985). *Applied statistical algorithm.* Ellis Horwood Limited

[4] Henze, N. (1986). A probabilistic representation of the 'skew-normal' distribution, *Scandinavian Journal of Statistics,* **13***, 271-275.*

[5] Hwang, J. T. and Casella, G. (1982). Minimax confidence sets for the mean of a multivariate normal distribution, *Annals of Statistics,* **10***, 868-881.*

[6] Hwang, J. T. and Casella, G. (1984). Improved set estimators for a multivariate normal mean, *Statistics and Decisions*, Supplement Issue, **1, 3-16**.

[7] Lewis, P. A. W. and Orav, E. J. (1989). *Simulation methodology for statisticians, operations analysts, and engineers Vol.* **1**, Wadsworth & Brooks/Cole Advanced Book & Software, Pacific Grove, California.

[8] Matin, M. A. (2005). Small-sample properties of estimators and tests in logistic regression model with skew-normally distributed explanatory variables, *Mauscript.*

[9] Matin, M. A. and Saleh, A. K. Md. E. (2005). Some improved estimators in Logistic regression model, *Journal of Statistical Research, Vol.* **39***, N0. 2, 37-58.*

[10] Morgan, B. J. T. (1984). *Elements of simulation.* Chapman and Hall.

[11] Robert, C. and Saleh, A. K. Md. E. (1989). Recentered confidence sets: An overview, *Technical Report*, Laboratorie de statistique theorique et applique, Universite Pierre et Marie Cuuie.

[12] Saleh, A. K. Md. E. and Sen, P. K. (1987). Estimation of regression parameters under uncertain prior restriction in a subspace. In *Recent development in statistics and actuarial science, pp. 171-187*. New York, SCTEX Publications.

[13] Wichmann, B. A. and Hill, I. D. (1982). Algorithm AS 183: An efficient and portable pseudo-random number generator, *Applied Statistics* (Correction in *Applied Statistics,* **33***, 123, 1984),* **31***, 188-190.*