A new two stage adaptive nonparametric test for paired differences

Weiwen Miao* and Joseph Gastwirth†

This paper proposes a new adaptive procedure for analyzing paired data. The procedure uses a function of ordered absolute values of the differences to measure tail heaviness of the underlying distribution. The value of the measure is then used to choose an appropriate signed rank test. The new adaptive procedure is shown to preserve the size of the test at its nominal level for all continuous distributions and typically has nearly the same power as the best signed rank test for a wide range of distributions.

Keywords and phrases: adaptive paired test, signed rank test, tail-heaviness.

1. MOTIVATION AND INTRODUCTION

Paired data arises in a wide variety of applications (e.g., Sprent and Smeeton, 2001). In a recent security law case where both authors were statistical consultants to the defendant lawyers, a Wall Street firm was charged of favoring a few customers who “shared profits” with them. Since the “profit sharing” funds usually sold their shares through a different firm than the one they acquired them, the regulators need to estimated the profit. Unfortunately, two authorities, the SEC (Securities and Exchange Commission) and the NASD (National Association of Securities Dealers) used different methods to estimate the profits customers made. A natural question is whether the two estimates of profits are essentially the same. As a firm could be examined by each regulator and be accused of “profit sharing” by one regulator and exculpated by the other. Since the two different measurements are for the same stock, the data is naturally paired. Data of 157 IPOs (Initial Public Offering) during an 18-month period were made available to us. Our analysis shows that the differences have heavier tail than the normal distribution. It’s better to use the $t_2$ or Cauchy scores test to detect the difference.

Paired data also arise in environmental studies where duplicate observations are taken. In 2003, the Michigan Department of Environmental Quality Remediation and Redevelopment Division conducted off-site samplings for the Detroit lead assessment project. Twenty-four soil samples were collected for lead analysis in vicinity of the former Great Lakes Smelting Company in Detroit, Michigan. These 24 samples are from 12 different locations near the plant with two samples from each location. Our interest is to test whether the two measurements from the same location are the same. Analysis shows that the differences are not normal. (The p-value of the Shapiro-Wilk test is 0.029.) Hence, the paired t-test is not appropriate for the analysis of this data.

Motivated by these two examples, we develop an adaptive procedure that uses a preliminary test to first analyze the tail-heaviness of the underlying distribution of the differences. That information is then used to choose an appropriate signed rank test to analyze the paired differences. For the sample of 157 estimated profits, our adaptive procedure chooses the $t_2$ scores test which yields a p-value of 0.009, about one-third of the one obtained from the paired t-test (p-value = 0.027). For the environmental data, the new adaptive procedure also chooses the $t_2$ scores test which yields a p-value of 0.038, showing that the two measurements are actually significantly different. For comparison purposes, the p-values for the t-test and the Wilcoxon test are 0.46 and 0.11, respectively. This example illustrates that in small sample sizes, the new adaptive procedure is able to detect a significant difference when both the t-test and the Wilcoxon test could not.

Typically, when the paired differences follow a normal distribution, the t-test is used; and when the normality assumption fails, the nonparametric Wilcoxon test is recommended. However, Freidlin, Miao and Gastwirth (2003) showed that this approach does not have high power when the paired differences are heavy-tailed. Randles and Hogg (1973) introduced an adaptive procedure that strictly maintains the size of the test and has high power when the distribution is heavy-tailed. Unfortunately, their procedure does not have high power for light to moderate tailed distributions. Freidlin et al. (2003) proposed an adaptive procedure (denoted by FMG) that uses the p-values of the Shapiro-Wilk normality test to choose an appropriate signed rank test to analyze the pairs. Their adaptive procedure has high power for a reasonable large class of moderate to heavy tailed distributions. However, their preliminary test is only asymptotically uncorrelated with the signed rank test used to analyze the paired sample, and the type I error for the procedure is slightly inflated. The FMG test was motivated by an environmental law case with sample size 16 and designed for...
relatively small samples. When the sample size is large, the
Shapiro-Wilk and other tests of normality will detect minor
deviations from normality. This may lead the FMG pro-
cedure to select a test that is good for a very heavy tailed
distribution even if the differences are not that far from normal.
The new method relying on a measure of “heavy-tailness”
orther than the p-value of a preliminary test is applicable in
large samples.

Mathematically, let \(X_1, \ldots, X_n\) be i.i.d. paired differences
from a continuous distribution \(F\). The \(X_i\) are naturally sym-
metric about a center \(\mu\). We are interested in testing whether
this center is equal to a known value \(\mu_0\). Without loss of gen-
erality, assume that \(\mu_0 = 0\), i.e. we are interested in testing:

\[
\begin{align*}
H_0 : \mu &= 0 \\
H_a : \mu &\neq 0.
\end{align*}
\]

It’s well known that when the \(X_i\)s follow a normal dis-
tribution, the optimal test is the t-test (Lehmann, 1986).
However, if we know that \(X\) comes from a heavy-tailed dis-
bution, e.g. \(\xi_2\) or Cauchy, then the signed rank tests with
\(\xi_2\) (Gastwirth, 1970) or Cauchy scores (Capon, 1961) have
higher power than the t-test. We propose an adaptive pro-
cedure (denoted by MG) that first uses functions of order
higher power than the t-test. We propose an adaptive pro-

procedure (denoted by MG) that first uses functions of order
higher power than the t-test. We propose an adaptive pro-

those are compared with the paired t-test, Wilcoxon test and the
FMG test.

2. PROPERTIES OF THE PRELIMINARY
TESTS

2.1 The preliminary tests

We want an adaptive procedure to have high power over
a wide range of distributions, from the light tailed uniform
to the heavy-tailed Cauchy distributions. The tests used
to analyze data at the second stage are signed rank tests.
Those tests are based on the absolute ranks. It’s well known
(Lemma 8.3.11, Randles and Wolfe, 1979) that the ranks
and the order statistics of the absolute values of the \(X_i\) are
independent. We choose functions of the ordered \(|X_i|\) as the
preliminary test, to measure the tail-heaviness. Doing so,
the preliminary test and the tests used to analyze data are
independent, which guarantees that the overall size of the
adaptive procedure is kept at the nominal level.

The most commonly used measure of spread is the sample
standard deviation, \(s\). However, some other statistics, such
as the median absolute deviation from the median are more
robust measures of the spread. Under the null hypothesis of
symmetry about 0, the corresponding sample statistics are:

\[
s^2 = \frac{1}{n} \sum_{i=1}^{n} X_i^2 \quad \text{and} \quad \tilde{M} = \frac{\text{median}(|X_1|, \ldots, |X_n|)}{\Phi^{-1}(0.75)},
\]

where \(\Phi(x)\) is the c.d.f. of standard normal. Note that the
constant \(1/\Phi^{-1}(0.75)\) in \(\tilde{M}\) is chosen to make \(\tilde{M}\) a consistent
estimator for the standard deviation when the underlying
distribution is normal.

When the data are normally distributed, the ratio of two
statistics, namely,

\[
sM = \frac{s}{\tilde{M}}
\]

is near 1. They will be greater (or less) than 1.0 when data
has heavier (or lighter) tails than normal. The statistic \(sM\)
provides us indication of heaviness of the tails. We then use
it as the preliminary test to choose the appropriate signed
rank tests to analyze the paired data.

If one estimates the center by the sample median rather
than the assumed 0, the statistic \(\tilde{M}\) becomes the median
absolute deviation from the median, which is asymptotically
equivalent to the semi-interquartile range, a robust estimate
of spread (Hall and Welsh, 1985).

2.2 Asymptotic properties of the preliminary

test

The following theorem shows that the statistic \(sM\) is
asymptotically normally distributed.

**Theorem 2.1.** Let \(X\) be a symmetric continuous ran-
dom variable with distribution function \(F\). Let \(f\) and \(\sigma^2\) be
its density function and the variance. Let \(F(0) = \frac{1}{2}\) and
\(\xi = F^{-1}(0.75)\), i.e. 0 and \(\xi\) are the median of \(X\) and \(|X|\),
respectively. Let

\[
\gamma_1 = \int_0^\infty xf(x)\,dx, \quad \gamma_2 = \int_0^\infty x^2f(x)\,dx,
\]

\[
\gamma_3 = \int_0^\infty x^3f(x)\,dx, \quad \gamma_4 = \int_0^\infty x^4f(x)\,dx.
\]

If the distribution \(F\) satisfies the conditions (A)–(C)
listed in the Appendix, then statistic \(sM\) is asymptotically
normal. In other words,

\[
\sqrt{n} \left[ sM - \frac{\sigma F^{-1}(0.75)}{\xi} \right] \Rightarrow N(0, \sigma^2_M),
\]
expressed as:  

$$s_{M} = \Phi^{-2}(0.75)$$

$$\times \left[ \frac{2\gamma_{4} - \sigma^{4}}{4\varepsilon^{2}\sigma^{2}} + \frac{\sigma^{2}}{16\xi^{4}f^{2}(\xi)} + \frac{\int_{-\infty}^{\xi} x^{2}f(x)dx - \frac{1}{\xi}\sigma^{2}}{2\xi^{3}f(\xi)} \right].$$

The proof of the Theorem is given in Appendix.

Table 1 lists the means and standard deviations of the asymptotic distributions of the $s_{M}$ statistic for some commonly used symmetric distributions.

### 3. THE ADAPTIVE PROCEDURES

#### 3.1 The signed rank tests

The signed rank tests used to analyze paired samples are expressed as:

$$S = \sum_{i=1}^{n} a(R_{i}^{+}\delta(X_{i}), \text{ with } a(i) = J\left(\frac{i}{n+1}\right)$$

where $R_{i}^{+}$ is the rank of $|X_{i}|$ among $|X_{1}|, |X_{2}|, \ldots, |X_{n}|$,

$$\delta(X_{i}) = \begin{cases} 1 & \text{if } X_{i} > 0 \\ 0 & \text{otherwise} \end{cases}$$

and $J$ is a score function appropriate for the data.

When the data comes from a light-tailed normal distributions or short-tailed uniform distributions, the normal scores test is known to have high power; when the tail-heaviness of the underlying distributions is somewhat medium, like logistic, double exponential or contaminated normal distributions, the Wilcoxon test is highly correlated with the maximin efficiency robust test (Gastwirth, 1966) and the Wilcoxon test should be used. Furthermore, if the data is heavy-tailed, e.g. from a $t_{2}$ or Cauchy distribution, then the appropriate signed rank test is the $t_{2}$ scores or the Cauchy scores test. We choose the following 4 score functions which include the extreme members of the $t$ family of distributions:

- $J_{1}(u) = \Phi^{-1}\left(\frac{1}{2} + \frac{1}{2}u\right)$, (Normal scores)
- $J_{2}(u) = u$, (Wilcoxon)
- $J_{3}(u) = \frac{2\sqrt{2\pi}}{\tan(\frac{\pi}{2}u)}u\sqrt{1 - u^{2}}$, (t2 scores)
- $J_{4}(u) = \frac{2\tan(\frac{\pi}{2}u)}{1 + \tan(\frac{\pi}{2}u)}$, (Cauchy scores)

#### 3.2 Percentiles of the preliminary tests and adaptive procedures

In order to form the cut-off of the adaptive procedure, we looked at both the empirical and asymptotic percentiles for statistic $s_{M}$ for different distributions. Table 2 lists the empirical percentiles for statistic $s_{M}$, for sample size 100. The distributions in the tables are arranged according to their tail heaviness, from the very heavy-tailed Cauchy to the light-tailed normal and uniform distributions. In the tables, C. Norm(5) (or C. Norm(3)) represents Tukey’s 10% contaminated normal 0.9N(0, 1) + 0.1N(0, 5²) (or 0.9N(0, 1) + 0.1N(0, 3²)). The results are based on $10^{5}$ simulations.

The general idea of the adaptive procedure is to use the sampling distribution of the preliminary test on a variety of underlying distributions to create cut-off values to classify the sample as light, medium, heavy or very heavy tailed. Then one chooses the signed rank test that is known to have high power for a distribution with tails of that type. We start with the heavy tailed Cauchy distribution, then move down the tail-heaviness to the light-tailed normal distribution, and finally to the short-tailed uniform distribution. Table 2 shows that more than 95% of Cauchy samples have $s_{M}$ values bigger than 2.7 ($5^{th}$ percentile), and more than 80% of $t_{2}$ samples have $s_{M} < 2.7$. Thus 2.7 is chosen as the cut-off value of $s_{M}$ between the Cauchy scores test and the $t_{2}$ scores test. In other words, if $s_{M} \geq 2.7$, the Cauchy scores test is used. If $s_{M}$ is slightly less than 2.7, the $t_{2}$ scores test should be used. Furthermore, Table 2 indicates that more than 97.5% $t_{2}$ samples, more than 90% contaminated normal samples (0.9N(0, 1) + 0.1N(0, 5²)) and more than 80% double exponential samples have $s_{M}$ values larger than 1.2 ($2.5^{th}$, $10^{th}$ and $20^{th}$ percentiles, respectively), while about 80% of logistic samples have $s_{M}$ values smaller than 1.2 ($80^{th}$ empirical percentile). Therefore we choose cut-off 1.2 between the $t_{2}$ scores test and the Wilcoxon test. This means that if $s_{M}$ is between 1.2 and 2.7, the $t_{2}$ scores test is used. Similarly, we choose 1.02 as the cut-off between the Wilcoxon test and the normal scores test. In other words, if $s_{M}$ is between 1.02 and 1.2, the Wilcoxon test is used, and if $s_{M} < 1.02$, the normal scores test is used. The cut off 1.02 was chosen because for sample size 100, all the uniform samples have $s_{M}$ values smaller than 1, but for sample size 25, about 95% of uniform samples have $s_{M}$ values less than 1.02. Furthermore, the Wilcoxon test has high power on both normal and logistic distributions. Consequently, the new adaptive procedure, denoted by MG, is the following:

- the MG Adaptive Test:
  - $s_{M} \geq 2.7$, use the Cauchy scores test;
  - $2.7 > s_{M} \geq 1.2$, use the $t_{2}$ scores test;
  - $1.2 > s_{M} \geq 1.02$, use the Wilcoxon test;
  - $s_{M} < 1.02$, use the Normal scores test.
4. SIZE AND POWER OF THE NEW ADAPTIVE PROCEDURE

4.1 Size

The preliminary test $sM$ is a function of ordered absolute values of the differences, and the tests used to analyze data are the signed rank tests based on absolute ranks. For continuous distributions, the signed ranks and the absolute order statistics are independent (Lemma 8.3.11 of Randles continuous distributions, the signed ranks and the absolute data are the signed rank tests based on absolute values of the differences, and the tests used to analyze distributions. Furthermore, under the alternatives, the $sM$ statistic and the p-value of the chosen test are also related. Simulations show that for light-tailed (heavy-tailed) underlying distributions, smaller values of the $sM$ statistic are associated with smaller (larger) p-values. This is consistent with theory, as the expected values of the $sM$ statistic are small on light-tailed distributions and small values of the $sM$ lead to the selection of either the Normal scores test or the Wilcoxon test at the second stage. Those two tests are known to have high power for light-tailed normal and logistic distributions.

4.2 Power

The following graphs are the size and power comparisons for 3 different procedures: The adaptive procedure proposed by Freidlin, Miao and Gastwirth (2003), the optimal signed rank tests for the particular distribution (denoted by Best SRT) and the new MG adaptive procedure. For contaminated normal distributions ($0.9N(0,1) + 0.1N(0,3^2)$), the Wilcoxon test is considered to be highly efficient (Hodges and Lehmann, 1961), and for uniform distributions, the normal score test is taken to be the best signed rank test. The graphs are based on $10^4$ simulations for samples of size 100.

Figures 1 and 2 indicate that the two adaptive procedures, FMG and MG, have about the same power as the best signed rank tests for normal and logistic distributions. It’s interesting to note that for double exponential distributions (Figure 3) the two adaptive procedures MG and FMG have slightly higher power than the the best signed rank test (the sign test). For heavy-tailed $t_2$ and contaminated normal distributions ($0.9N(0,1) + 0.1N(0,3^2)$), the MG adaptive test has about the same power as the best signed rank test; their powers are slightly higher than the FMG procedure. For sample size 100, the power of the MG procedure is about 2-6% higher than the FMG test for contaminated normal distributions. But when the sample size increases to 200, the power of the MG test is significantly higher than the FMG test (about 10–15% higher). (Also see Table 3 below.) This occurs because in large samples, the Shapiro-Wilk test will yield a low p-value even when the

\begin{table}
\centering
\caption{Percentiles of the statistic $sM$ on a Variety of Distributions ($n=100$)}
\begin{tabular}{ccccccc}
\hline
\textbf{Symmetry} & \textbf{Cauchy} & \textbf{$t_2$} & \textbf{C. Norm(5)} & \textbf{D. Exp.} & \textbf{C. Norm(3)} & \textbf{Logistic} & \textbf{Normal} & \textbf{Uniform} \\
\hline
1% & 2.0944 & 1.1450 & 1.0218 & 1.0165 & 0.9089 & 0.8827 & 0.8237 & 0.8684 \\
2.5% & 2.3968 & 1.2135 & 1.0894 & 1.0587 & 0.9439 & 0.9106 & 0.8458 & 0.8695 \\
5% & 2.7533 & 1.2804 & 1.1563 & 1.0974 & 0.9789 & 0.9365 & 0.8661 & 0.7062 \\
10% & 3.2813 & 1.3696 & 1.2426 & 1.1461 & 1.0219 & 0.9688 & 0.8903 & 0.7197 \\
20% & 4.1983 & 1.4984 & 1.3588 & 1.2110 & 1.0803 & 1.0107 & 0.9230 & 0.7377 \\
25% & 4.6705 & 1.5564 & 1.4065 & 1.2372 & 1.1053 & 1.0278 & 0.9360 & 0.7450 \\
30% & 5.1890 & 1.6129 & 1.4504 & 1.2614 & 1.1282 & 1.0434 & 0.9481 & 0.7517 \\
40% & 6.3855 & 1.7309 & 1.5335 & 1.3079 & 1.1712 & 1.0736 & 0.9712 & 0.7646 \\
50% & 7.9517 & 1.8615 & 1.6167 & 1.3540 & 1.2140 & 1.1020 & 0.9934 & 0.7772 \\
60% & 10.219 & 2.0171 & 1.7036 & 1.4026 & 1.2607 & 1.1330 & 1.0169 & 0.7907 \\
70% & 13.914 & 2.2263 & 1.8008 & 1.4583 & 1.3133 & 1.1676 & 1.0439 & 0.8057 \\
75% & 16.798 & 2.3644 & 1.8554 & 1.4911 & 1.3439 & 1.1879 & 1.0591 & 0.8145 \\
80% & 21.114 & 2.5508 & 1.9191 & 1.5284 & 1.3787 & 1.2106 & 1.0768 & 0.8246 \\
90% & 43.597 & 3.2484 & 2.0970 & 1.6356 & 1.4768 & 1.2753 & 1.1265 & 0.8531 \\
95% & 87.661 & 4.2296 & 2.2534 & 1.7303 & 1.5655 & 1.3336 & 1.1703 & 0.8790 \\
97.5% & 173.56 & 5.6334 & 2.4012 & 1.8198 & 1.6458 & 1.3874 & 1.2116 & 0.9033 \\
99% & 436.05 & 8.5566 & 2.5813 & 1.9310 & 1.7480 & 1.4548 & 1.2657 & 0.9340 \\
\hline
\end{tabular}
\end{table}
Table 3. Power comparison for the FMG, MG and the Best SRT tests

<table>
<thead>
<tr>
<th>Distribution</th>
<th>Sample Size</th>
<th>Alternative</th>
<th>MG</th>
<th>FMG</th>
<th>Best SRT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal</td>
<td>25</td>
<td>0.627</td>
<td>0.8268</td>
<td>0.8377</td>
<td>0.8460</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>0.31</td>
<td>0.8565</td>
<td>0.8523</td>
<td>0.8700</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>0.215</td>
<td>0.8415</td>
<td>0.8324</td>
<td>0.8497</td>
</tr>
<tr>
<td>Logistic</td>
<td>25</td>
<td>1.11</td>
<td>0.8288</td>
<td>0.8487</td>
<td>0.8479</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>0.545</td>
<td>0.8554</td>
<td>0.8629</td>
<td>0.8691</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>0.374</td>
<td>0.8478</td>
<td>0.8379</td>
<td>0.8566</td>
</tr>
<tr>
<td>Double Exponential</td>
<td>25</td>
<td>0.805</td>
<td>0.8375</td>
<td>0.8577</td>
<td>0.8496</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>0.365</td>
<td>0.8733</td>
<td>0.8747</td>
<td>0.8558</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>0.25</td>
<td>0.8700</td>
<td>0.8652</td>
<td>0.8478</td>
</tr>
<tr>
<td>$t_2$</td>
<td>25</td>
<td>0.878</td>
<td>0.8134</td>
<td>0.829</td>
<td>0.8497</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>0.4</td>
<td>0.8432</td>
<td>0.8145</td>
<td>0.8524</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>0.275</td>
<td>0.8369</td>
<td>0.7997</td>
<td>0.8427</td>
</tr>
<tr>
<td>$0.9N(0,1) + 0.1N(0, 3^2)$</td>
<td>25</td>
<td>0.735</td>
<td>0.8374</td>
<td>0.8563</td>
<td>0.8512</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>0.35</td>
<td>0.8379</td>
<td>0.7703</td>
<td>0.8526</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>0.245</td>
<td>0.8316</td>
<td>0.6953</td>
<td>0.8557</td>
</tr>
</tbody>
</table>

Figure 1. Power comparison for normal distributions.

Figure 2. Power comparison for logistic distributions.

data are nearly normal, e.g. contaminated normal, and the FMG procedure will choose the $t_2$ or Cauchy scores test too often on contaminated normal distributions. For uniform distributions, again the MG adaptive test has about the same power as the normal scores test, but the FMG adaptive procedure has significantly lower power compared to the other two procedures. This is not surprising as the normal scores test, which is known to have high power on uniform distributions, is not one of the second stage choices in the FMG test.

The following table presents the powers of the FMG, the MG and the best signed rank test for sample sizes at 200, for Normal, Logistic, Double exponential, $t_2$ and contaminated Normal distributions. The powers of the tests for samples of size 25 and 100 are also included. The purpose of the table is to illuminate the roles of the sample sizes in determining the properties of the adaptive test. The alternatives are chosen so that the best signed rank test has about 85% power. The results are based on $10^4$ simulations. The table clearly shows that for $t_2$ and contaminated normal distributions, when the sample size gets larger, the new adaptive test $MG$ is better than the FMG test.

Adaptive test for paired data 217
Simulations were also performed on the Cauchy distribution for samples of size 25, 100 and 200. The powers of the FMG, the MG and the Cauchy scores tests are about the same.

The main reason that the adaptive procedure MG is powerful is that the preliminary test chooses either the best or the second best signed rank test most of the time. Simulations show that for normal distributions, our adaptive test chooses the normal scores test about 62% of the time, the Wilcoxon test about 35% of the time, and only chooses the
\[ t_2 \text{ scores test about 3\% of the time. For samples from a } t_2 \text{ distributions, the } MG \text{ procedure chooses the } t_2 \text{ scores test about 82\% of the time, the Cauchy scores test about 14\% of time, and the Wilcoxon test about 4\% of time.} \]

These power simulations indicate that the adaptive procedure MG keeps the size at the nominal level and generally has the same power as the best signed rank test for a variety of symmetric distributions. This procedure is recommended for use in practice.

5. EXAMPLES

5.1 The securities hypothetical profit data

This data concerns the hypothetical profit estimates for 157 IPOs during an 18-month period. The p-value of paired t-test is 0.027, indicating that the two measurements from two different regulators are statistically different. The data is clearly not normal (the p-value of the Shapiro-Wilk test is 0.029, so the differences are not normally distributed. We are interested in whether the two measurements from each of the 12 different locations in the Detroit area.

The authors would like to thank reviewers for several helpful suggestions.

APPENDIX

The following are the conditions (A)–(C) in Theorem 2.1:

(A) \( F^{-2}(x) \) is continuous on the open unit interval \((0,1)\) and satisfies a first order Lipschitz condition on every interval bounded away from 0 and 1; \( (F^{-2})'(x) \) exists and is continuous except on a set of Jordan content 0;

(B) \( f(x) \neq 0 \) and the Riemann integral

\[
\int_0^1 (F^{-2})'(z)/(2z+1) |u(1-u)|^{1/2} du \text{ converges absolutely;}
\]

(C) For \( i = 1, 2 \) there exist \( \delta_i \in (0,1) \) such that \( \forall k_i > 0 \exists M_i < \infty \) such that

\( M_1^{-1} < \frac{F^{-1}\left(\frac{u+1}{2}\right)(F^{-1})'(\frac{u+1}{2})}{F^{-1}\left(\frac{u+1}{2}\right)(F^{-1})'(\frac{u+1}{2})} < M_1; \)

\( M_2^{-1} < \frac{F^{-1}\left(\frac{u+1}{2}\right)(F^{-1})'(\frac{u+1}{2})}{F^{-1}\left(\frac{u+1}{2}\right)(F^{-1})'(\frac{u+1}{2})} < M_2. \)

\textit{Proof of Theorem 2.1.} As } X \text{ is symmetric about 0,

\[ f_{\mid X}\mid (x) = 2f(x), \quad F_{\mid X}\mid (x) = 2F(x) - 1, \text{ and } F_{\mid X}\mid^{-1}(u) = F^{-1}\left(\frac{u+1}{2}\right). \]
It can be shown that
\[
E(\hat{s}^2) = \sigma^2, \quad E(\hat{M}) = \frac{\xi}{\Phi^{-1}(0.75)},
\]
\[
\text{Var}(\hat{s}^2) = \frac{1}{n} [2\gamma_4 - \sigma^4],
\]
\[
\text{Var}(\hat{M}) = \frac{1}{16n f^2(\xi)[\Phi^{-1}(0.75)]^2}.
\]

Let
\[
M = \text{median}(|X_1|, \ldots, |X_n|), \quad I_i = \begin{cases} 1 & \text{if } |X_i| < \xi \\ 0 & \text{otherwise.} \end{cases}
\]

Consider the interval \((\xi, M)\). Note that this interval contains \(\frac{n}{2} - \sum I_i\) observations as \(M\) is the \(\frac{n+1}{2}\)th observation and \(\sum I_i\) is the # of observations less than \(\xi\). Using the Bahadur representation, \(f_{|X|}(\xi)(M - \xi) = 2f(\xi)(M - \xi) \approx P(|X| \in (\xi, M))\). Consequently,
\[
2nf(\xi)(M - \xi) \sim \frac{n}{2} - \sum I_i = -\sum_{i=1}^{n} (I_i - \frac{1}{2}).
\]

This implies that
\[
M - \xi \sim -\frac{1}{2nf(\xi)} \sum_{i=1}^{n} (I_i - \frac{1}{2}).
\]

Hence,
\[
\text{Cov}(\hat{s}, M) = \text{Cov}\left(\frac{1}{n} \sum_{i=1}^{n} X_i^2 - \frac{1}{2nf(\xi)} \sum_{i=1}^{n} (I_i - \frac{1}{2})\right)
\]
\[
= -\frac{n}{2nf(\xi)} \text{Cov}(X_i^2, I_i - \frac{1}{2})
\]
\[
= -\frac{1}{2nf(\xi)} \left[ E(X_i^2 I_i) - E(X_i^2)E(I_i) \right]
\]
\[
= -\frac{1}{2nf(\xi)} \left[ \int_{-\xi}^{\xi} \frac{\Phi^{-1}(0.75)}{\xi^4} \int_{-\xi}^{\xi} x^2 f(x)dx - \frac{1}{2}\sigma^2 \right]
\]

and
\[
\text{Cov}(\hat{s}, \hat{M}) = -\frac{1}{2nf(\xi)} \Phi^{-1}(0.75) \left[ \int_{-\xi}^{\xi} x^2 f(x)dx - \frac{1}{2}\sigma^2 \right].
\]

Let \(X_{[i]}\) be the order statistics of \(|X_i|\). Since both \(\hat{s}^2\) and \(\hat{M}\) are L-statistics, they can be represented as:
\[
\hat{s}^2 = \frac{1}{n} \sum_{i=1}^{n} c_i h(|X_{[i]}|) + a |X_{[\frac{n+1}{2}]}|
\]

with appropriate function \(h(\cdot)\) and constants \(c_i\) and \(a\), where \(\lceil \frac{n+1}{2} \rceil\) is the largest integer smaller than \(\frac{n+1}{2}\). The function \(h\) and constants \(c_i\) and \(a\) corresponding to the statistics \(\hat{s}^2\) and \(\hat{M}\) are:

\(\text{For } c^2, c = 1, h(x) = x^2; a = 0;\)
\(\text{For } \hat{M}, c_i = 0, h(x) = x^2; a = \Phi^{-1}(0.75).\)

The joint asymptotic normality of \((\hat{s}^2, \hat{M})\) is obtained by applying Corollary 3 in Chernoff et al. (1967). It can be checked that the conditions (A)–(C) listed in this Theorem are equivalent to the assumptions \(A^*, B^{**}\) and \(E\) in Chernoff et al. (1967) with their function \(H(u) = F^{-2}(\frac{u}{2})\). According to Remark 9 of Chernoff et al. (1967), asymptotically \((\hat{s}^2, \hat{M})\) is normal with mean \((\sigma^2, \frac{\xi}{\Phi^{-1}(0.75)})\).

The Taylor expansion of \(sM = \sqrt{\frac{s^2}{M}}\) at \((\sigma^2, \frac{\xi}{\Phi^{-1}(0.75)})\) is:
\[
sM = \sqrt{\frac{s^2}{M}} = \sqrt{\frac{\sigma^2}{\xi\Phi^{-1}(0.75)}} + \frac{1}{2\sigma \xi\Phi^{-1}(0.75)} (\hat{s}^2 - \sigma^2)
\]
\[
- \frac{\sqrt{\sigma^2}}{\xi\Phi^{-2}(0.75)} (M - \xi) + \frac{1}{2\sigma\xi} (\hat{s}^2 - \sigma^2)
\]
\[
= \frac{\sigma\Phi^{-1}(0.75)}{\xi} + \frac{\Phi^{-1}(0.75)}{\xi^2} (\hat{s}^2 - \sigma^2)
\]
\[
- \frac{\sigma\Phi^{-2}(0.75)}{\xi^2} (M - \xi) + \frac{1}{2\sigma\xi} (\hat{s}^2 - \sigma^2).
\]

Consequently,
\[
\sqrt{n}[sM - \frac{\sigma\Phi^{-1}(0.75)}{\xi}] \Rightarrow N(0, \sigma^2_M)
\]

with
\[
s^2_M = \frac{n\Phi^{-2}(0.75)}{4\xi^2\sigma^2} \text{Var}(\hat{s}^2) + \frac{na^2\Phi^{-4}(0.75)}{\xi^4} \text{Var}(\hat{M})
\]
\[
- \frac{n\Phi^{-3}(0.75)}{\xi^3} \text{Cov}(\hat{s}^2, \hat{M})
\]
\[
= \Phi^{-2}(0.75) \left[ \frac{2\gamma_4 - \sigma^4}{4\xi^2\sigma^2} + \frac{\sigma^2}{16\xi^4 f^2(\xi)} + \frac{\int_{-\xi}^{\xi} x^2 f(x)dx - \frac{1}{2}\sigma^2}{2\xi^2 f(\xi)} \right].
\]

\[\square\]

Received 11 November 2008

REFERENCES

Weiven Miao  
Department of Mathematics  
Haverford College  
Haverford, PA 19041.  
E-mail address: wmiao@haverford.edu

Joseph Gastwirth  
Department of Statistics  
George Washington University  
Washington, DC, 20052.  
E-mail address: jlgast@gwu.edu