An adaptive test for the two-sample scale problem where the common quantile may be different from the median

W. Kössler[a,*], Narinder Kumar[b]

[a] Institute of Computer Science, Humboldt-Universität zu Berlin, Germany
[b] Department of Statistics, Panjab University, Chandigarh, India

---

Abstract

In the usual two-sample scale problem it is assumed that the two populations have a common median. We consider the case where the common quantile may be other than a half. We investigate a quite general class, all members are based on U-statistics where the minima and maxima of subsamples of various sizes are used. The asymptotic efficacies are investigated in detail. We construct an adaptive test where all statistics involved are suitably chosen. It is shown that the proposed adaptive test has good asymptotic and finite power properties.

Keywords: Adaptive test, asymptotic efficacy, scale alternative, U-statistics, tailweight

---

1. Introduction

Let $X_1, \ldots, X_{n_1}$ and $Y_1, \ldots, Y_{n_2}$ be independent random samples from a population with absolutely continuous distribution functions $F(x)$ and $F(x/e^{\vartheta})$. We wish to test

$$H_0: \quad \vartheta = 0$$

against the alternative

$$H_1: \quad \vartheta \neq 0.$$

The general two-sample scale problem was considered by Kochar and Gupta (1986), Kössler (1994, 1999), Hall et. al. (1997), Ramsey and Ramsey (2007), and more recently, by Kössler and Kumar (2010) and Marozzi (2011, 2012).

* Corresponding Author, Rudower Chaussee 25, D-12489 Berlin, Tel.:+49 3020933087

Preprint submitted to Statistical Methodology *August 17, 2015*

Here, we consider the scale case when the populations have a common known quantile that may be of an order other than half. Such problems have many practical applications. As an example, cf. Deshpande and Kusum (1984) , let us have two filling machines that shall fill half of kg cans of dried milk. According to a given laid down criterion not more than five percent of the cans be under filled. Therefore both the machines are adjusted in such a way that five percent of the cans contain less than half kg and 95 percents can contain more than half kg of dried milk. In this case we can say that the distribution of the amount filled by the machines have a common quantile of order 0.05 and the more efficient machine is the one with smaller dispersion around this quantile.

Another field where such problem may occur is the pharmaceutical industry. Liu et.al. (2013) report on a drug-drug interaction study on a rheumatoid arthritis test drug where the test compound is taken by patients which are already under a particular medication (MTX, in this case). However the test compounds may result in a delayed elimination of MTX which is toxic if it remains too long in the body. Therefore the 24h MTX plasma level should not exceed a certain threshold value t, say 5picograms, with given probability q, e.q. $q = 0.9$. Assume now we have two test drugs under study, both satisfy the quantile condition $F(t) = q$ for the MTX level. Then the question is whether the MTX levels of the two test drugs differ in scale.

The ranked-set setting of such a question was considered in Öztürk and Deshpande (2004) and in Gaur et al. (2013).

Assume without restriction to the generality the common quantile is zero, i.e. $F(0) = \alpha$. Mehra and Rao (1992) suggested the follwing kernel

$$\Phi(x_1, \ldots, x_k, y_1, \ldots, y_k) = \begin{cases} 1 & \text{if} \quad 0 \leq x_{(k)k} < y_{(k)k} \text{ and } 0 \leq x_{(1)k}, y_{(1)k} \\ 1 & \text{if} \quad y_{(1)k} < x_{(1)k} < 0 \text{ and } x_{(k)k}, y_{(k)k} < 0 \\ -1 & \text{if} \quad 0 \leq y_{(k)k} < x_{(k)k} \text{ and } 0 \leq x_{(1)k}, y_{(1)k} \\ -1 & \text{if} \quad x_{(1)k} < y_{(1)k} < 0 \text{ and } x_{(k)k}, y_{(k)k} < 0 \\ 0 & \text{otherwise,} \end{cases}$$

where $x_{(i)k}$ is the ith order statistic in a subsample of size k from the X-sample (and likewise for y's). The suggested test statistic is

$$U_k = \frac{n_1 n_2}{\binom{n_1}{k} \cdot \binom{n_2}{k}} \sum \Phi(X_{r_1}, \ldots, X_{r_k}, Y_{s_1}, \ldots, Y_{s_k}),$$

where the summation extends over all possible combinations $(r_1, \ldots, r_k)$ of k

integers from $\{1,\ldots,n_1\}$ and all possible combinations $(s_1,\ldots,s_k)$ of k integers from $\{1,\ldots,n_2\}$. Obviously, large values of U_k speak for more scaled Y-sample.

Note that the kernel presented is a natural extension of Despande and Kusum (1984) [1] and Kusum (1985) [12].

The article is organized as follows. In Section 2 we give a rank representation of the test statistic that simplifies their computation considerably. In Section 3 we compute their asymptotic efficacies and intend to determine a suitable choice of subsample size k w.r.t. tail behaviour of the underlying distribution. The results are used to define an adaptive test in Section 4. A simulation study is performed in Section 5. Section 6 gives a short example of the application of our test. Conclusions are drawn in Section 7.

2. Rank representation of U_k

Let us consider the positive and negative observations separately and let $n_j^<$ and $n_j^\geq$, $j=1,2$ the numbers of the negative and nonnegative observations, respectively. Let $R_{(s)}$ be the rank of the $Y_{(s)}$ observation in the pooled sample and $Q_{(s)} = R_{(s)} - n_1^< - n_2^<$. Assume first that we have no tied observations.

Then we may write U_k as

$$U_k = \frac{n_1 n_2}{\binom{n_1}{k}\cdot\binom{n_2}{k}}\Big((T_{1k}+T_{2k})-(T_{3k}+T_{4k})\Big),$$

with

$$\begin{aligned}
T_{1k} &= \;=\sum_{s=k}^{n_2^\geq}\binom{s-1}{k-1}\binom{Q_{(s)}-s}{k} \\
T_{3k} &= \binom{n_1^\geq}{k}\binom{n_2^\geq}{k} - T_{1k} \\
T_{4k} &= \sum_{s=1}^{n_2^<}\sum_{j=1}^{k}\binom{n_2^<-s}{k-1}\binom{R_{(s)}-s}{j}\binom{n_1^<+s-R_{(s)}}{k-j} \\
T_{2k} &= \binom{n_1^<}{k}\binom{n_2^<}{k} - T_{4k}
\end{aligned}$$

where the rank representations of the components are derived from Kumar, et al. (2003), cf. also Kössler and Kumar (2010). Note that Mehra and Rao (1992) obtained a different rank representation of U_k.

There are several opportunities to handle ties, cf. e.g. Hajek et.al. (1999). We prefer to assign random ranks to the tied observations. This procedure is equivalent to add a very small random error to the tied observations and it has the advantage that the (asymptotic) distribution properties of U_k are preserved.

3. Asymptotic distributions and asymptotic efficacies

Let $N = n_1 + n_2$. The asymptotic distribution of $\sqrt{N}(U_k - \mathbf{E}(U_k))$ as $N \to \infty$ in such a way that $n_1/N \to \lambda$, $0 < \lambda < 1$, is normal with mean zero and variance

$$\sigma_k^2 = \frac{k^2 \eta_{10}}{\lambda} + \frac{k^2 \eta_{01}}{1-\lambda}$$

where

$$\begin{aligned} \eta_{10} &= \text{var}(\phi_{1,0}(X)) \\ \eta_{01} &= \text{var}(\phi_{0,1}(X)) \\ \phi_{1,0}(x) &= \mathbf{E}(\Phi(x, X_2, \ldots, X_k, Y_1, \ldots, Y_k) \\ \phi_{0,1}(x) &= \mathbf{E}(\Phi(X_1, \ldots, X_k, y, Y_2, \ldots, Y_k). \end{aligned}$$

Under H_0 we see that $\mathbf{E}(U_k) = 0$ and

$$\sigma_k^2 = \frac{k^2}{(4k-1)\lambda(1-\lambda)} \left(\alpha^{4k-1} + (1-\alpha)^{4k-1} \right),$$

cf. Mehra and Rao (1992) [16]. Moreover, under the sequence of alternatives $\theta_N = \theta N^{-1/2}$ the asymptotic efficacy (AE_k) of U_k is

$$AE_k = \frac{\eta_k^2}{\sigma_k^2}$$

where

$$\eta_k = \lambda(1-\lambda) \cdot 2k^2 \int_{-\infty}^{\infty} |x|(F(x)-\alpha)^{2k-2} f^2(x)\, dx$$

For $k = 1$ the statistic U_k reduces to Despande and Kusum (1984)[1] and for $k = 2$ it reduces to Kusum (1985) [12].

In the following we provide some AE_k (except for the factor $\lambda(1-\lambda)$) for some values of k, various densities, and for $\alpha = 0.5, 0.1, 0.05, 0.01$. That AE_k

that are the largest for a given density and a given α value are written in bold style.

We see from Table 1 that for the uniform we have the largest AE_k for k as large as possible, which is not surprising. For the other densities we always have optimal values for some k, all k are less than or equal to 4. Moreover, for densities with shorter tails (normal, exponential) the choice of $k = 3$ or $k = 4$ is the best, for densities with medium tails (logistic, Gumbel) the choice of $k = 2$ is the best, for that with longer tails (DE, t_2) $k = 1$ (or $k = 2$, dependent on α) should be chosen. To confirm these findings we performed further calculations, for various t-densities, especially for the Cauchy, cf. Table 2.

4. An adaptive test

Considering our original examples, values of α about 0.05 or 0.1 are of most interest here, and the calculations of the previous section suggest to introduce the following adaptive test. We apply the concept of Hogg (1974) , which was already applied for the ordinary two-sample scale problem by Kössler (1994), Hall et al. (1997) and, more recently by Marozzi (2012). That is, we classify at first the type of the underlying density with respect to one measure of tailweight t, which is defined by

$$t \quad = \quad \frac{F^{-1}(0.95) - F^{-1}(0.05)}{F^{-1}(0.85) - F^{-1}(0.15)}. \tag{1}$$

An estimate $\hat{t}$, is obtained by inserting $\hat{Q}(u)$, the so-called classical quantile estimate of $F^{-1}(u)$in (1), with

$$\hat{Q}(u) = \begin{cases} X_{(1)} - (1-\delta) \cdot (X_{(2)} - X_{(1)}) & \text{if} \quad u < 1/(2 \cdot N) \\ (1-\delta) \cdot X_{(j)} + \delta \cdot X_{(j+1)} & \text{if} \quad \frac{1}{2 \cdot N} \leq u \leq \frac{2 \cdot N - 1}{2 \cdot N} \\ X_{(N)} + \delta (X_{(N)} - X_{(N-1)}) & \text{if} \quad u > (2 \cdot N - 1)/(2 \cdot N), \end{cases} \tag{2}$$

where $\delta = N \cdot u + 1/2 - j$ and $j = \lfloor N \cdot u + 1/2 \rfloor$.

Define regions $E_1, \ldots, E_4$ which separate the space of continuous distributions into four disjunct subsets, $E_1 = \{t > 2.0\}$ (long tails), $E_2 = \{1.6 \leq t \leq 2.0\}$ (medium tails), $E_3 = \{1.4 \leq t < 1.6\}$ (short tails), and $E_4 = \{t < 1.4\}$ (very short tails).

The cutoff values of the regions are determined in such a way that the vast majority of densities is classified correctly, i.e. they fall in the class that has the

Table 1: AE_k (except for the factor $\lambda(1-\lambda)$) for $k = 1, \ldots, 8$, uniform, normal, Gumbel, exponential, logistic, DE, t_2 distribution

density	tailw.	α	k=1	k=2	k=3	k=4	k=5	k=5	k=7	k=8
uniform	1.286	0.50	3	7	11	15	19	23	27	**31**
		0.10	2.76	6.30	9.90	13.5	17.1	20.7	24.3	**27.9**
		0.05	2.87	6.65	10.45	14.25	18.05	21.85	25.45	**29.45**
		0.01	2.97	6.93	10.89	14.85	18.81	22.77	26.73	**30.69**
normal	1.587	0.50	1.22	1.68	1.82	**1.85**	1.84	1.81	1.78	1.75
		0.10	2.21	3.05	3.27	**3.28**	3.25	3.15	3.08	2.99
		0.05	3.03	4.13	**4.35**	4.34	4.24	4.13	4.00	3.87
		0.01	5.33	6.81	**6.96**	6.81	6.57	6.32	6.01	5.85
Gumbel	1.655	0.50	1.23	1.72	1.87	**1.92**	1.91	1.89	1.86	1.83
		0.10	1.31	**1.56**	1.54	1.47	1.40	1.32	1.26	1.20
		0.05	1.65	**1.94**	1.89	1.78	1.68	1.58	1.50	1.42
		0.01	2.50	**2.73**	2.57	2.38	2.20	2.05	1.93	1.81
exponen.	1.697	0.50	2.36	4.95	7.31	9.55	11.73	13.82	15.98	**18.06**
		0.10	0.71	0.82	**0.83**	0.81	0.79	0.76	0.73	0.70
		0.05	0.72	0.87	**0.88**	0.86	0.83	0.80	0.77	0.74
		0.01	0.74	0.90	**0.92**	0.90	0.86	0.83	0.80	0.77
logistic	1.697	0.50	1.05	1.33	**1.36**	1.33	1.28	1.22	1.18	1.13
		0.10	2.27	**2.85**	2.83	2.69		2.38	2.24	2.12
		0.05	3.39	**4.18**	4.08	3.82	3.56	3.32	3.11	2.92
		0.01	7.26	**8.31**	7.77	7.08	6.45	5.94		5.11
DE	1.912	0.50	0.75	0.91	**0.93**	0.90	0.87	0.84	0.81	0.78
		0.10	2.72	**2.78**	2.34	2.06	1.83	1.67	1.54	1.44
		0.05	4.66	**4.73**	3.92	3.30	2.87	2.33	2.23	2.14
		0.01	**11.83**	11.52	9.09	7.33	6.16	5.35	4.75	4.29
t_2	2.107	0.50	0.75	**0.78**	0.69	0.60	0.53	0.47	0.42	0.38
		0.10	2.59	**2.63**	2.18	1.79	1.49	1.26	1.09	0.95
		0.05	**5.19**	5.11	4.09	3.26	2.66	2.21	1.87	1.61
		0.01	**26.0**	23.4	17.5	13.3	10.3	8.28	6.79	5.69

Table 2: AE_k (except for the factor $\lambda(1-\lambda)$) for $k = 1, 2, 3$, various t_ν densities

density	tailweight	α	k=1	k=2	k=3
t_1	3.217	0.50	**0.49**	0.41	0.31
		0.10	**4.00**	3.08	1.98
		0.05	**14.15**	10.77	6.60
		0.01	**317.2**	237.6	138.8
t_2	2.107	0.50	0.75	**0.78**	0.69
		0.10	2.59	**2.63**	2.18
		0.05	**5.19**	5.11	4.09
		0.01	**26.0**	23.4	17.5
t_3	1.798	0.50	0.88	**0.99**	0.93
		0.10	2.38	**2.67**	2.42
		0.05	4.11	**4.49**	3.95
		0.01	**13.5**	13.4	11.1
t_4	1.786	0.50	0.95	**1.13**	1.10
		0.10	2.31	**2.73**	2.58
		0.05	3.73	**4.30**	3.97
		0.01	10.1	**10.7**	9.36
t_5	1.737	0.50	1.00	**1.22**	1.21
		0.10	2.28	**2.78**	2.69
		0.05	3.54	**4.23**	4.00
		0.01	8.69	**9.52**	8.60

highest asymptotic power (cf. Tables 1 and 2). For example, the normal (tailweight t=1.59) is classified to E_3, and the test U_3, which is the best among the considered tests, is performed. The logistic is mapped to region E_2 and the 1test U_2 is performed. Similar observations for the other densities lead to the given cutoff values. In few cases, if the classification doesn't be correct, then the efficacy loss is very small in almost all cases.

Now, we propose the simple Adaptive test A which is based on the four U-statistics U_k, $k = 1, 2, 3, 4$ and on the selector statistic $S = \hat{t}$,

$$\text{choose } U_k \quad \text{if} \quad S \in E_k.$$

Since the Adaptive test A is based on the concept of Hogg (1974) it is asymptotically distribution-free, and the asymptotic power of the Adaptive test A is that of the test U_k when the underlying density belongs to region E_k, cf. e.g. Kössler (2006).

5. Simulation study

In order to assess whether the asymptotic theory can also be applied for medium to small sample sizes a simulation study (10,000 replications each for the null case, 1,000 replications each for the alternative cases) is performed. We choose the following eight distributions:

- Uniform distribution (density with very small tailweight),
- Normal distribution (density with small tailweight),
- Exponential distribution (skew density with small tailweight),
- Logistic distribution (density with medium tailweight),
- Doubleexponential distribution (density with large tailweight),
- Cauchy distribution (density with very large tailweight),
- Gumbel distribution (skew density),
- a scale contaminated normal (density with medium tailweight)

We consider the four single U-tests U_k, and the Adaptive test $A(\hat{t})$.

The sample sizes $n_1 = n_2 = 10, 25, 50, 100, 400$ as well as $n_1 = 100, n_2 = 200$ and $n_1 = 200, n_2 = 100$ (the latter only for $\alpha = 0.05$) and the alternatives $\theta_N = N^{-1/2}\theta$ with various θ are considered. Estimated levels of significance are summarized in Table 3 for the uniform density. For the other densities we get similar values except for the adaptive test. For the adaptive test the estimated level is similar to that for the best test for the respective density. E.g. for the uniform

Table 3: Estimated levels of significance for the various tests, common α quantile, uniform density

α	$n_1 = n_2$	U_1	U_2	U_3	U_4	Adaptive
0.1	25	0.0538	0.0582	0.0678	0.0765	0.0753
	50	0.0488	0.0527	0.0599	0.0648	0.0637
	100	0.0512	0.0521	0.0542	0.0585	0.0585
	400	0.0489	0.0478	0.0513	0.0507	0.0507
0.05	25	0.0509	0.0575	0.0615	0.0693	0.0686
	50	0.0509	0.0516	0.0561	0.0600	0.0592
	100	0.0498	0.0473	0.0522	0.0536	0.0537
	400	0.0534	0.0521	0.0505	0.0523	0.0523
	n_1=100, n_2=200	0.0489	0.0496	0.0516	0.0539	0.0538
	n_1=200, n_2=100	0.0510	0.0522	0.0522	0.0537	0.0537

the adaptive test satisfies the level if and only if the test U_4 satisfies it, for the normal adaptive test satisfies the level if and only if the test U_3 satisfies it.

For $n_1 = n_2 = 400$ the level is satisfied for all tests, for $n_1, n_2 \geq 100$ this fact is true for U_1, U_2 and perhaps for U_3. U_4 is slightly anticonservative, for $n_1 = n_2 = 50$ U_3 is also slightly anticonservative. For $n_1 = n_2 = 25$ all tests are more or less anticonservative, where the degree of anticonservativity of U_i increases with the index i. Generally, the degree of anticonservativity of U_i increases with the index i and decreases with the sample size.

The power is as expected, the adaptive test is always the best or the second best (after the asymptotically best) among the four single tests U_k, $k = 1, 2, 3, 4$. This assertion is true for all sample sizes.

To get an impression, for $n_1 = n_2 = 400$ the results of the simulation study are summarized in Figure 1. For some densities, such as for the normal or contaminated normal the power curves of various tests look very similar.

6. Illustrative example

To illustrate our procedure we consider a factory that produces 500ml milk packets. The packages are filled in such a way that they consist at least the critical value of 500ml with probability of at least 95%. Assume, there are two machines of maybe different costs. The management of the factory has to decide which of the two machines is to use. They prefer that machine with significantly lower

Figure 1: Power functions for $m = n = 400$ of the various U-tests and for the adaptive test.

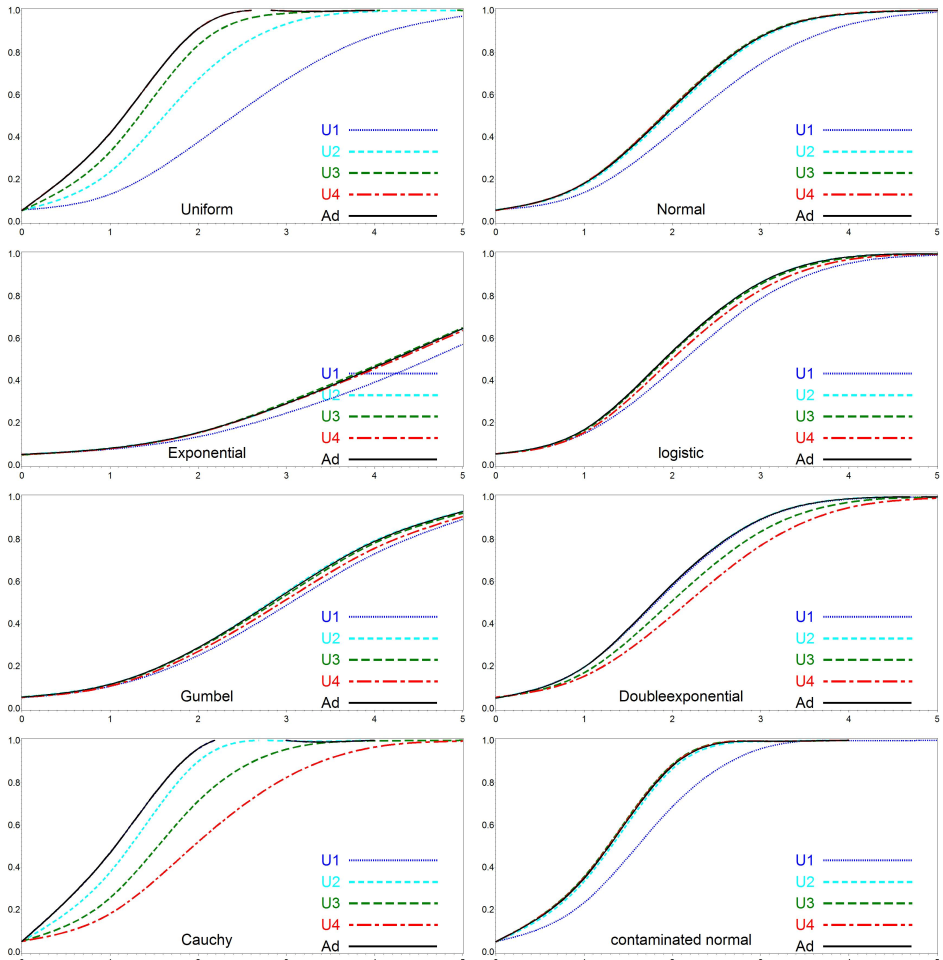
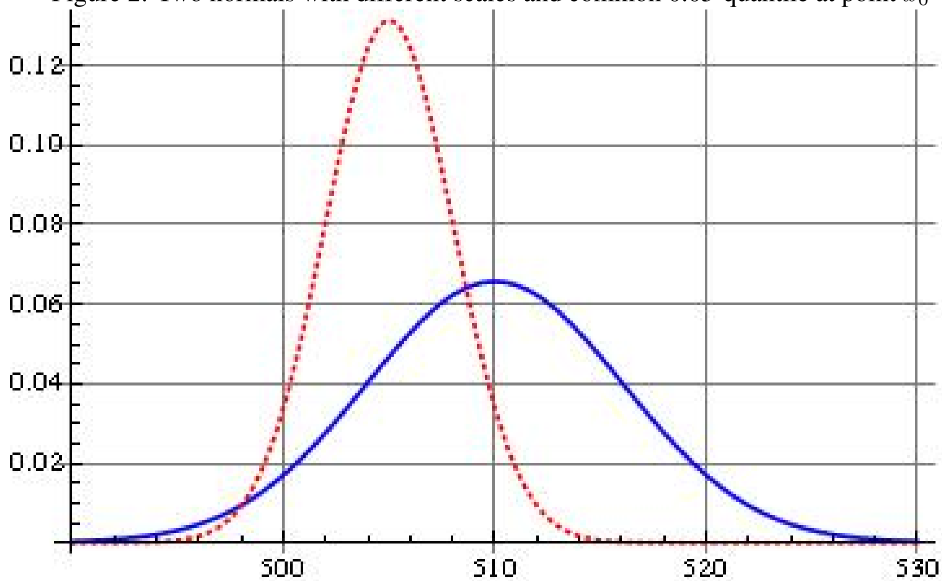


Figure 2: Two normals with different scales and common 0.05-quantile at point $x_0 = 500$



variance as in this case less milk is used.

In our simulation we used two normals, each with 0.05-quantile at point 500, $X \sim \mathcal{N}(505, 3.04^2)$ and $Y \sim \mathcal{N}(510, 6.08^2)$ (cf. Figure 2) In our example we obtain $U_1 = 0.511, U_2 = 0.646, U_3 = 0.649, U_4 = 0.613$ and the adaptive test chooses test U_2 which is asymptotically almost as good as the best test. All p-values are very small, $p < 0.0001$.

7. Conclusions

The asymptotics works well for sample sizes from 400 on. That seems to be rather large, however, as we use small quantiles as small as 0.01, 0.05 and 0.1, that result might be expected. Note that we also, for smaller sample sizes, tried to estimate the variance or to use permutation methods, but the results did not turn better.

The results of our study may be summarized as follows.

- use single or adaptive test if $n_1 = n_2 \geq 400$
- if $i = 1, 2$ then use single or adaptive test if $n_1 = n_2 \geq 100$
- if there are long tails take k=1, for medium tails take $k = 2$ or $k = 3$, and for short tails take $k = 4$ (or, perhaps, more)

- if density is unknown use the adaptive test

Acknowledgement

The authors thank the referees whose comments led to an improvement of the paper. We also thank Wolf F. Lesener for his assistance in writing various SAS programs.

References

[1] Deshpande, J.V., Kusum, K., 1984. A test for the nonparametric two-sample scale problem. Australian Journal of Statistics 26, 16–24.

[2] Gaur, A., Mahajan, K.K., Arora, S., 2013. A nonparametric test for a multi-sample scale problem using ranked-set data. Statistical Methodology 10, 85–92.

[3] Hájek, J., Sidák, Z., Sen, P.K., 1999. Theory of Rank tests. Wiley New York.

[4] Hall, P., Padmanabhan, A., 1997. Adaptive inference for the two-sample scale problem. Technometrics 39, 412–422.

[5] Hogg, R., 1974. Adaptive robust procedures: partial review and some suggestions for future applications and theory. Journal of the American Statistical Association 69, 909–923.

[6] Kochar, S.C., Gupta, R., 1986. Some competitors of the mood test for the two-sample scale problem. Comm.Statist. Theor.Meth. 15(1), 231–239.

[7] Kössler, W., 1994. Restrictive adaptive tests for the treatment of the two-sample scale problem. Computational Statistics & Data Analysis 18, 513–524.

[8] Kössler, W., 1999. Rank tests in the two-sample scale problem with unequal and unknown locations. Statistical Papers 40, 13–36.

[9] Kössler, W., 2006. Restrictive adaptive tests for restricted alternatives. Shaker Aachen.

[10] Kössler, W., Kumar, N., 2010. An adaptive test for the two-sample scale problem based on U-statistics. Communications in Statistics - Computation and Simulation 39, 1785–1802.

[11] Kumar, N., Singh, R., Öztürk, Ö., 2003. A new class of distribution-free tests for location parameters. Sequential Analysis 22, 107–128.

[12] Kusum, K., 1985. A new distribution-free test for two-sample scale problem. Journal of the Indian Statistical Association 23, 97–101.

[13] Liu, W., Bretz, F., Hayter, A.J., Glimm, E., 2013. Simultaneous inference for several quantiles of a normal population with spplications. Biometrical Journal 55, 360–369.

[14] Marozzi, M., 2011. Levene type tests for the ratio of two scales. Journal of Statistical Computation and Simulation 81, 815–826.

[15] Marozzi, M., 2012. A modified Hall-Padmanabhan test for the homogeneity of scales. Communications in Statistics - Theory and Methods 41, 3068–3078.

[16] Mehra, K.L., Rao, K.S., 1992. A class of two-sample nonparametric tests for scale based on sub-sample extrema, in: 80th Indian Science Congress conference held at Goa University, Goa, India.

[17] Öztürk, O., Deshpande, J.V., 2004. A new nonparametric test using ranked set data for a two-sample scale problem. Sankhya Ser A 66, 513–527.

[18] Ramsey, P.H., Ramsey, P.P., 2007. Testing variability in the two-sample case. Communications in Statistics - Simulation and Computation 36, 233–248.