THE MOST EFFICIENT LINEAR COMBINATION OF THE SIGN AND THE MAESONO TESTS FOR *P*-NORMAL DISTRIBUTIONS

G. Burgio, S. Patrì

1. INTRODUCTION

In (Burgio and Nikitin, 2001) the linear combination $G = aS + bW$ of the Sign test S and the Wilcoxon test W was studied. The Pitman efficiency of G with respect to Student's t, in the case of location parameters, and its maximum with respect to a and b were calculated and it was obtained that S and W are locally most powerful and Pitman optimal signed rank tests for the Laplace and the logistic densities. In this paper, we obtain the most efficient G combined test when the parent distribution belongs to the p-normal family of densities

$$f_p(x;\theta,\sigma_p) = [2p^{1/p}\Gamma(1+1/p)\sigma_p]^{-1}\exp(-|x-\theta|^p / p\sigma_p^p)$$

$$\sigma_p = \sigma p^{-1/p}[\Gamma(1/p)/\Gamma(3/p)]^{1/2}$$

where $x \in \Re$, θ and $\sigma > 0$ are real parameters and p is the structural parameter ranging between 0 and ∞.

The most relevant cases regard the range $(1,\infty)$, to which we will limit our consideration in this paper. Relevant cases of the p-normal density are: for $p=1$ the double exponential or Laplace distribution, for $p=2$ the normal density, and for $p\to\infty$ the rectangular density $(2\sigma\sqrt{3})^{-1}$ with range $\theta - \sigma\sqrt{3} < x \leq \theta + \sigma\sqrt{3}$.

The Pitman asymptotic efficiency of the Sign test, for $p=1$ reaches 2 (the highest value), for $p=2$ is $2/\pi$ and for $p\to\infty$ tends to 1/3. The Sign test is asymptotically more efficient than the t test for $1\leq p\leq 1,4$ i.e. for p-normal distributions with kurtosis $4\leq\beta_2\leq 6$. The Wilcoxon test (Burgio, 1996) is asymptotically more efficient than the t test for $1\leq p\leq 1,75$ i.e. for p-normal distributions with kurtosis $3,3\leq\beta_2\leq 6$.

The paper proceeds as follows.

In Section 2, we present the properties of the G combined test and show that, for any p, it is more efficient of both the Sign and the Wilcoxon tests.

In Section 3, we consider the properties of the Maesono test G_4 and compare it with the G test, which is a particular case of the Maesono's one for $r=2$.

In Section 4, we may conclude that G_2 for leptokurtic p-normal distributions, and of G_4 for platikurtic p-normal distributions are, in general, more efficient than Student's t, with maximum loss of efficiency of 3,1% in the near proximity of the normal distribution.

2. PROPERTIES OF THE G COMBINED TEST

In (Burgio and Nikitin, 2001) some properties of G have been proved. If we consider that G is a sequence of U statistics (Hoeffding, 1948) with kernel

$$\Phi_{a,b}(s,t) = a(1_{\{s>0\}} + 1_{\{t>0\}})/2 + b1_{\{s+t>0\}}$$

it can be represented in the form

$$G = \binom{n}{2} \sum_{1 \le i < j \le n} \Phi_{a,b}(X_i, X_j)$$

where X_i (i=1,n) are the observed values of a random sample of size n.

Using the central limit theorem for U-statistics (Hoeffding, 1948), G is asymptotically normally distributed. For weakly unimodal symmetric densities (Hodges and Lehmann, 1956), the Pitman relative asymptotic efficiency of the G combined test is ≥ 1 with respect to both S and W.

Suppose that, under the null-hypothesis of symmetry H_0, the initial distribution function (d.f) $F_0(x)$, is absolutely continuous and symmetric with respect to zero. Hence, for every x, 1- $F_0(x)$- $F_0(-x)$=0.

Under the alternative hypothesis H_1, the observations have common d.f. $F(x,\theta)$, θ>0, such that $F(x,\theta)= F_0(x)$ for some symmetric d.f. F_0, with continuous density f_0 only for θ=0.

Therefore, the case of location alternative is $F(x,\theta)= F_0(x-\theta)$.

The Pitman relative asymptotic efficiency of G with respect to t is

$$e(G;t) = \frac{12\sigma^2 \left[af_0(0) + 2b \int_{-\infty}^{+\infty} f_0^2(y)dy \right]^2}{3a^2 + 6ab + 4b^2}$$

which, for the standard normal density, becomes

$$e(G;t) = \frac{6(a + b\sqrt{2})^2}{\pi(3a^2 + 6ab + 4b^2)}.$$

The max Pitman efficiency of G relative to t is 0,9643, higher than 0,6366 for the Sign test and 0,9549 for the Wilcoxon test individually considered.

In our work, we found that the combined G test has an efficiency, relative to t, which equals

$$e(G;t,p)=\frac{(a+2^{1-1/p}b)^2 3p^2\Gamma(3/p)}{(3a^2+6ab+4b^2)\Gamma^3(1/p)}.$$

After some algebra, the maximal Pitman efficiency of the combined G test, for the p-normal family is

$$\max e(G;t,p)=\frac{p^2\Gamma(3/p)}{\Gamma^3(1/p)}[3(2^{1-1/p}-1)^2+1]$$

and the Pitman most efficient combination of the Sign and Wilcoxon tests for p-normal distributions is

$$G=\frac{4-3(2^{1-1/p})}{3(2^{1-1/p}-1)}S+W\ ,\ \forall p>1.$$

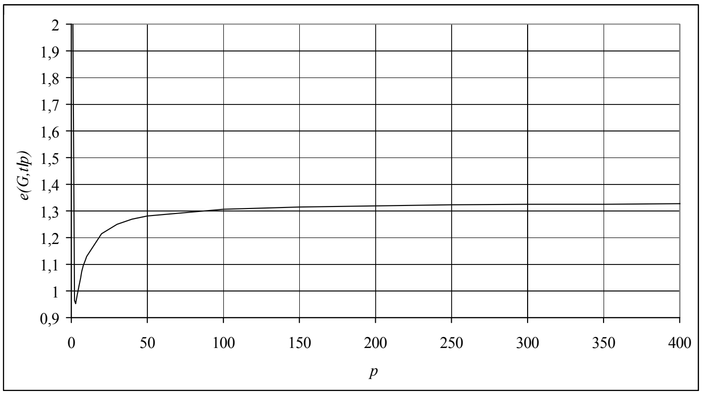
For $p\rightarrow 1$, the Pitman most efficient G test is the Sign test which has efficiency double than that of Student's t.

For $p=ln2/ln1{,}5\approx 1{,}7095$, $e(G;\,t,p)$ = 1,01153.

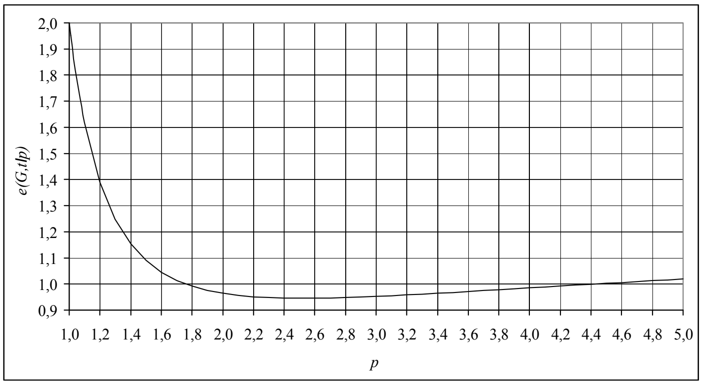
For $p\rightarrow\infty$ the most efficient combination is $G = W - 2/3\ S$ with Pitman efficiency, relative to t, equal to 4/3.

From *Graph 1b*, it is possible to note that the G test is less efficient than t, i.e. $e(G;\,t,p)<1$, for $1{,}76<p<4{,}4$ a range corresponding to a large interval of platikurtic and leptokurtic distributions with kurtosis index $2{,}1334<\beta_2<3{,}288$.

However, from *Graph 1a*, it is evident that, for platikurtic distributions with $p\geq 4{,}4$ (i.e. with kurtosis index $1{,}8\leq\beta_2\leq 2{,}13$), the G test is more efficient than Student's t.

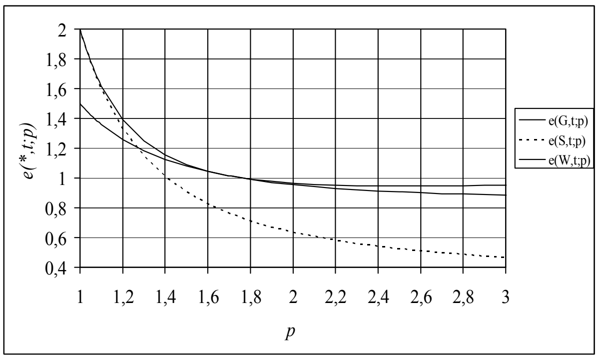


Graph 1a – Maximum Pitman efficiency of G (relative to t), as a function of $1\leq p\leq 400$.

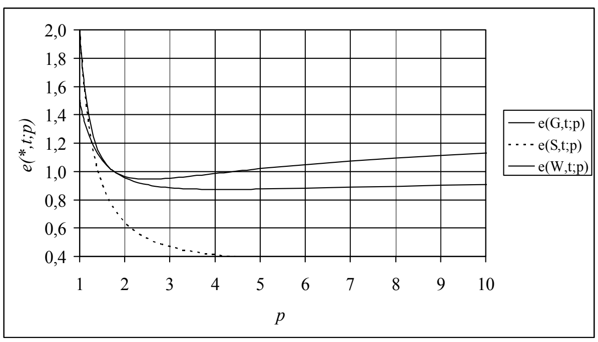


Graph 1b – Maximum Pitman efficiency of *G* (relative to *t*), as a function of 1≤ *p*≤5.

From the following *Graphs 2a* and *2b*, we can see that the combined test *G* is, almost everywhere, more efficient than its components *S* and *W*. In Table 1, the values of the *a* coefficients (*b*=1) providing for the most efficient *G* combination, with respect to *t*, are given.



Graph 2a – Pitman efficiency of *S*, *W* and the best *G* tests (relative to *t*), as a function of 1≤ *p*≤3.



Graph 2b – Pitman efficiency of *S*, *W* and the best *G* tests (relative to *t*), as a function of 1≤ *p*≤10.

TABLE 1

Values of a (b=1) maximising the Pitman efficiency e(G; t, p)

p	*a*	*e(G; t, p)*	*p*	*a*	*e(G; t, p)*
1	*a*=1; *b*=0	2	3,6	-0,48696	0,97123
1,01	47,40426	1,95144	3,7	-0,49367	0,97466
1,02	23,35953	1,90562	3,8	-0,49989	0,97813
1,04	11,33743	1,82148	4,0	-0,51109	0,98510
1,05	8,93312	1,78281	4,1	-0,51615	0,98858
1,07	6,18547	1,71156	4,3	-0,52533	0,99551
1,1	4,12497	1,61797	4,6	-0,53719	1,00571
1,2	1,72193	1,39083	4,7	-0,54071	1,00904
1,3	0,92167	1,24780	4,8	-0,54405	1,01233
1,4	0,52198	1,15374	4,9	-0,54721	1,01557
1,5	0,28244	1,08984	5,0	-0,55022	1,01878
1,6	0,12294	1,04540	6	-0,57363	1,04837
1,7	0,00915	1,01400	7	-0,58921	1,07360
1,8	-0,07610	0,99161	8	-0,60032	1,09506
1,9	-0,14233	0,97563	9	-0,60865	1,11342
2,0	-0,19526	0,96430	10	-0,61512	1,12925
2,1	-0,23852	0,95641	20	-0,64230	1,21532
2,2	-0,27454	0,95111	30	-0,65071	1,25048
2,3	-0,30499	0,94778	40	-0,65481	1,26953
2,4	-0,33107	0,94597	50	-0,65723	1,28146
2,5	-0,35365	0,94534	100	-0,66200	1,30653
2,6	-0,37339	0,94563	150	-0,66356	1,31526
2,7	-0,39080	0,94664	200	-0,66434	1,31970
2,8	-0,40627	0,94823	250	-0,66481	1,32239
2,9	-0,42009	0,95027	300	-0,66512	1,32420
3,0	-0,43253	0,95267	350	-0,66534	1,32549
3,1	-0,44377	0,95536	400	-0,66551	1,32646
3,2	-0,45399	0,95826	450	-0,66564	1,32722
3,3	-0,46331	0,96133	500	-0,66574	1,32783
3,4	-0,47186	0,96454	1000	-0,66620	1,33057
3,5	-0,47971	0,96785	∞	-2/3	4/3

3. PITMAN EFFICIENCY OF THE LINEAR COMBINATION OF *S* AND THE MAESONO TEST

In Burgio and Nikitin (2003) we considered a new test $G_r = aS + bW_r$, which is a linear combination of the Sign test S and the Maesono (1987) test W_r.

For r=2, the Maesono statistic coincides with the Wilcoxon test and G coincides with G_2. The Maesono statistic, for any natural $r\geq 2$, is a U-statistic

$$W_r = \binom{n}{r}^{-1} \sum_{1\leq i_1<\ldots<i_r\leq n} K_r(X_{i_1},\ldots,X_{i_r})$$

with the kernel

$$K_r(s_1,\ldots,s_r) = r^{-1}\left(\sum_{i=1}^{r}\prod_{j\neq i}^{r} 1_{(s_i+s_j>0)}\right).$$

The case r=3 is not interesting because G_3 has the same Pitman efficiency as G_2 (Maesono, 1987).

The G_r test has Pitman efficiency (Burgio and Nikitin, 2003)

$$e(G_r, f_0) = \frac{\left[af_0(0) + 2(r-1)\int F_0^{r-2}(x) f_0^2(x)dx \right]^2}{\frac{a^2}{4} + 2\left[\frac{1}{2r-1} - \frac{(r-1)!^2}{(2r-1)!}\right] + \frac{2a}{r}(1-2^{-r+1})} \tag{1}$$

whose maximum value, with respect to *a*, is given by

$$a_0 = \frac{2dk - 4mw}{2md - k} \tag{2}$$

where

$$d = 2(1-2^{-r+1})/r$$

$$k = 2(r-1)\int_{-\infty}^{+\infty} F_0^{r-2}(x) f_0^2(x)dx$$

$$m = f_0(0)$$

$$w = 2\left(\frac{1}{2r-1} - \frac{(r-1)!^2}{(2r-1)!}\right).$$

The maximum value of the Pitman efficiency (1), for $a=a_0$, is given by

$$e(G_r, f_0 \mid a = a_0) = \frac{k^2 - 4mdk + 4m^2 w}{w - d^2}. \tag{3}$$

The statistic G_4, with the appropriate choice of *a*, when the distribution is normal, has Pitman efficiency, relative to Student's *t*, 0,9794, higher than that of G_2, which is 0,9643.

In (Burgio and Nikitin, 2003) it is also shown that, for the logistic distribution, the Pitman efficiency of G_4 relative to *t* is equal to 1,0939.

As we may only compare the generalised test with other tests whose efficiency is known, we evaluated the Pitman most efficient combination of G_4 for the *p*-normal family

$$f(z_p) = [2p^{1/p}\Gamma(1+p^{-1})]^{-1}\exp(-\left|z_p\right|^p / p) \tag{4}$$

which, for $z_p>0$, has *df*

$$F(z_p) = 1/2 + \frac{1}{2\Gamma(1/p)}\gamma(1/p, z_p^p)$$

where $\gamma\left(\frac{1}{p}, z_p^p\right)$ is the Euler's incomplete Gamma function.

Unfortunately, for the p-normal family, it is not possible to get an explicit expression of the integral

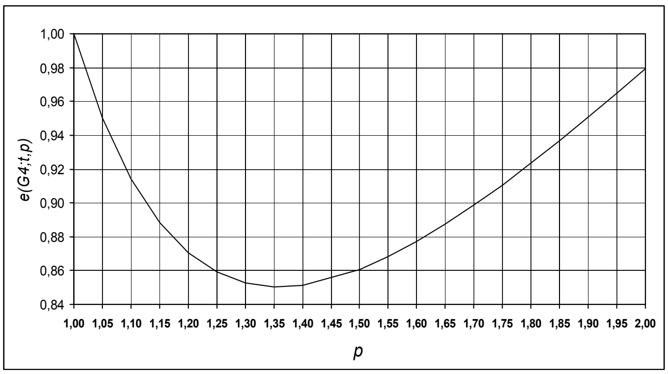
$$\int_{-\infty}^{+\infty} F_0^2(x) f_0^2(x) dx \tag{5}$$

which is part of the G_r Pitman efficiency (1).

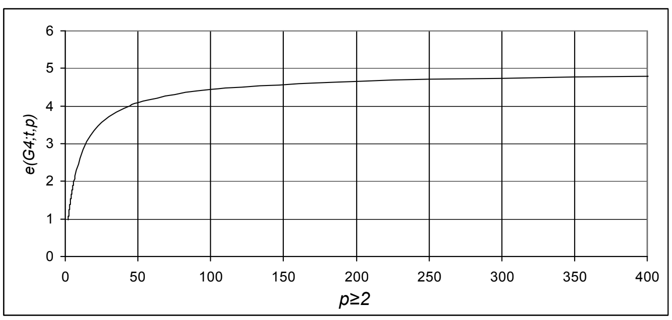
Therefore, we evaluated the integral (5) numerically and calculated a_0 and $e(G_r, f_0 \mid a = a_0)$ with the formulas (2) and (3), for r=4 and $p \geq 1$.

The G_4 combinations of the Sign and the Maesono W_4 tests with the highest Pitman efficiency relative to t, for the p-normal family ($p \geq 1$) are shown in Table 2.

The maximum Pitman efficiency of G_4 (relative to t) is shown in *Graph 3a* for $1 \leq p \leq 2$ and in *Graph 3b* for $2 \leq p \leq 400$.



Graph 3a – Maximum Pitman efficiency of G_4 (relative to t) for $1 \leq p \leq 2$.



Graph 3b – Maximum Pitman efficiency of G_4 (relative to t) for $2 \leq p \leq 400$.

TABLE 2

Values of a (b=1) maximising the Pitman efficiency $e(G_4; t, p)$

p	a	$e(G_4; t, p)$	p	a	$e(G_4; t, p)$	p	a	$e(G_4; t, p)$	p	a	$e(G_4; t, p)$
1,00	∞	1	2,45	-0,25308	1,11812	3,90	-0,41931	1,54373	5,35	-0,47521	1,88798
1,05	9,05397	0,95037	2,50	-0,26414	1,13381	3,95	-0,42214	1,55695	5,40	-0,47649	1,89853
1,10	4,23196	0,91425	2,55	-0,27448	1,14948	4,00	-0,42488	1,57007	5,45	-0,47773	1,90899
1,15	2,62484	0,88846	2,60	-0,28418	1,16513	4,05	-0,42753	1,58308	5,50	-0,47896	1,91938
1,20	1,82156	0,87072	2,65	-0,29329	1,18074	4,10	-0,43009	1,59599	5,55	-0,48015	1,92969
1,25	1,33971	0,85929	2,70	-0,30187	1,19629	4,15	-0,43257	1,60880	5,60	-0,48132	1,93992
1,30	1,01860	0,85289	2,75	-0,30996	1,21179	4,20	-0,43498	1,62151	5,65	-0,48247	1,95008
1,35	0,78929	0,85051	2,80	-0,31760	1,22723	4,25	-0,43732	1,63412	5,70	-0,48358	1,96016
1,40	0,61738	0,85138	2,85	-0,32483	1,24260	4,30	-0,43958	1,64663	5,75	-0,48468	1,97016
1,45	0,48370	0,85491	2,90	-0,33168	1,25789	4,35	-0,44178	1,65904	5,80	-0,48576	1,98009
1,50	0,37678	0,86061	2,95	-0,33818	1,27310	4,40	-0,44391	1,67136	5,85	-0,48681	1,98995
1,55	0,28933	0,86810	3,00	-0,34436	1,28822	4,45	-0,44598	1,68358	5,90	-0,48784	1,99974
1,60	0,21646	0,87707	3,05	-0,35024	1,30326	4,50	-0,44799	1,69571	5,95	-0,48885	2,00945
1,65	0,15483	0,88726	3,10	-0,35584	1,31821	4,55	-0,44995	1,70774	6	-0,48984	2,01909
1,70	0,10200	0,89848	3,15	-0,36118	1,33306	4,60	-0,45186	1,71967	10	-0,53379	2,61380
1,75	0,05622	0,91054	3,20	-0,36628	1,34782	4,65	-0,45371	1,73151	15	-0,55372	3,06499
1,80	0,01618	0,92331	3,25	-0,37115	1,36248	4,70	-0,45551	1,74326	20	-0,56326	3,35785
1,85	-0,01915	0,93667	3,30	-0,37581	1,37704	4,75	-0,45726	1,75492	25	-0,56887	3,56511
1,90	-0,05056	0,95052	3,35	-0,38028	1,39150	4,80	-0,45897	1,76649	30	-0,57255	3,72046
1,95	-0,07866	0,96478	3,40	-0,38456	1,40585	4,85	-0,46064	1,77796	35	-0,57517	3,84173
2,00	-0,10395	0,97937	3,45	-0,38867	1,42011	4,90	-0,46226	1,78935	40	-0,57711	3,93934
2,05	-0,12682	0,99424	3,50	-0,39261	1,43426	4,95	-0,46384	1,80065	45	-0,57862	4,01980
2,10	-0,14762	1,00932	3,55	-0,39640	1,44831	5,00	-0,46539	1,81186	50	-0,57982	4,08738
2,15	-0,16661	1,02459	3,60	-0,40005	1,46225	5,05	-0,46689	1,82299	100	-0,58518	4,43780
2,20	-0,18402	1,04000	3,65	-0,40356	1,47609	5,10	-0,46836	1,83403	200	-0,58784	4,65637
2,25	-0,20004	1,05551	3,70	-0,40694	1,48983	5,15	-0,46980	1,84499	300	-0,58872	4,74069
2,30	-0,21483	1,07110	3,75	-0,41020	1,50346	5,20	-0,47120	1,85586	400	-0,58916	4,78624
2,35	-0,22852	1,08675	3,80	-0,41334	1,51699	5,25	-0,47257	1,86665	600	-0,58960	4,83508
2,40	-0,24124	1,10243	3,85	-0,41638	1,53041	5,30	-0,47390	1,87736	∞	-62/105	1184/239

4. CONCLUSIONS

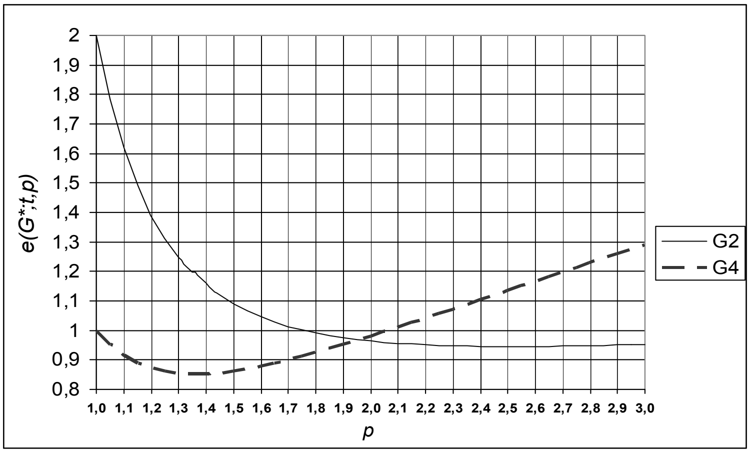
We can now easily compare the Pitman efficiency of the best G_2 and G_4 tests, with respect to the *t*-test, by varying the *p* structural parameter of the *p*-normal family. As we can see from the *Graphs 4a* and *4b*, Student's *t* appears to be a bit more efficient than G_2 and G_4 only in the proximities of the normal density (p=2) and in particular for $1{,}76<p<2{,}07$, that is for *p*-normal densities with kurtosis index $2{,}93<\beta_2<3{,}29$.

By using either the G_2 or the G_4 tests, instead of the Student's *t*, the maximum absolute loss of efficiency is about 3,1% when p=1,96 corresponding to a *p*-normal density with kurtosis index 3,04.

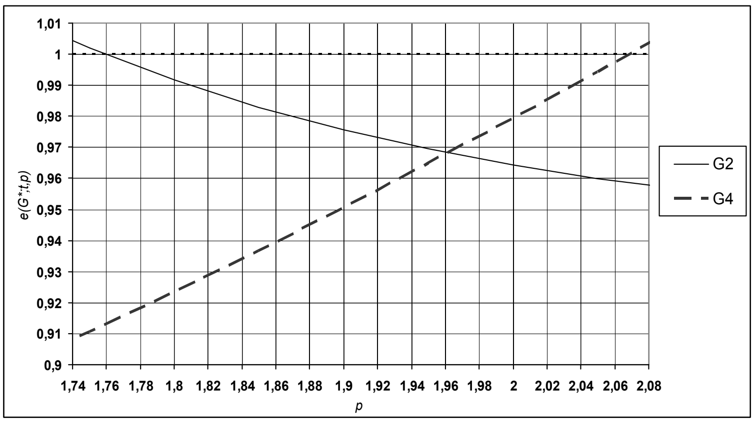
For values of *p* between 1 (double exponential density) and 1,76 (*p*-normal with kurtosis index 2,93) the G_2 test is much more efficient than the *t*-test, with maximum efficiency twice that of the Student's *t* for the double exponential density.

For platikurtic densities, and in particular for values of *p* between 2,07 and ∞ (rectangular density), the G_4 test is increasingly more efficient than the *t*-test. In particular, for the rectangular density, G_4 is about five times more efficient than Student's *t*.

It is therefore possible to conclude that the G_2 test, for leptokurtic *p*-normal distributions, and the G_4 test, for platikurtic *p*-normal distributions, are in general more efficient than Student's *t*, with a maximum loss of efficiency of about 3,1% in the near proximity of the normal distribution.



Graph 4a – Pitman efficiency of the best G_2 and G_4 tests (relative to *t*) as a function of $1\leq p\leq 3$.



Graph 4b – Pitman efficiency of the best G_2 and G_4 tests (relative to *t*) for $1{,}74\leq p\leq 2{,}08$.

5. AKNOWLEDGEMENTS

The present paper has been written jointly by the two Authors. Evaluations in paragraph 3 are due to Stefano Patrì.

Department of Methods and Models for Economics, Territory and Finance
Sapienza University of Rome

GIUSEPPE BURGIO
STEFANO PATRÌ

REFERENCES

G. BURGIO, YA. YU. NIKITIN (2001), *The combination of the Sign and Wilcoxon tests for symmetry and their Pitman efficiency*, "Asymptotic Methods in Probability and Mathematical Statistics", 42, pp. 12-34.

G. BURGIO, YA. YU. NIKITIN (2003), *On the combination of the Sign and Maesono tests for symmetry and its efficiency*, "Statistica", 63, 2, pp. 213-222.

J. HODGES AND E. LEHMANN (1956), *The efficiency of some non parametric competitors of the t-test*, "Annals of Mathematical Statistics", 27, pp. 324-335.

W. HOEFFDING (1948), *A class of statistics with asymptotically normal distributions*, "Annals of Mathematical Statistics", 19, pp. 293-325.

Y. MAESONO (1987), *Competitors of the Wilcoxon signed rank test*, "Annals of the Institute of Statistical Mathematics", 39, Pt A, pp. 363-375.

SUMMARY

The most efficient linear combination of the Sign and the Maesono tests for p-normal distributions

This paper deals with the Pitman most efficient G_r linear combination of the Sign and the Maesono tests for parent distributions belonging to the *p*-normal family of densities.

The most efficient linear combinations G_2 and G_4 are obtained. It is also shown that G_2 (for leptokurtic *p*-normal distributions) and G_4 (for platikurtic *p*-normal distributions) are much more efficient than Student's *t*, with a maximum loss of efficiency of about 3,1% in the near proximity of the normal distribution (p=2).