Metodološki zvezki, Vol. 4, No. 1, 2007, 9-20

Indices of Skewness Derived from a Set of Symmetric Quantiles: A Statistical Outline with an Application to National Data of E.U. Countries

Maurizio Brizzi[1]

Abstract

In this paper, which follows a recent field of research started by Tukey (1977), a class of indices of skewness is introduced, based on a symmetric set of quantiles. Two kinds of normaisation are proposed, leading to different indices, called VCS (Ventile Coefficient of Skewness) and VIS (Ventile Index of Slewness), respectively. The sample distribution of both indices is studied by a Monte Carlo simulation. Two extended indices of skewness (ECS and EIS) are proposed, having interesting inferential properties. Finally, an application to national data of 27 E.U. countries is presented, with a brief comment of the results.

.

1 Introducing the problem

The most known and successful index of skewness ever proposed is surely Pearson's γ, defined as the ratio of the third central moment to the cube of standard deviation. However, the most recent research lines about skewness do follow a quantile pattern. Such an approach, having the aim to define robust, efficient and user-friendly indices of shape (skewness and kurtosis) has been followed by several Authors, such as Tukey (1977), Antille *et al.* (1982), Hoaglin *et al.* (1985), Mac Gillivray (1986), Kappenman (1988), Arnold and Groeneveld (1995), Groeneveld (1998), Wang and Serfling (2005). In two recent papers (Brizzi, 2000 and 2002), we proposed and studied a class of indices of shape (skewness and kurtosis) based on *letter values*, which are symmetric quantiles whose analysis gives a particular stress to tails. In the present study, we do propose a class of indices of skewness which are built by taking into account all the sample body, the center as well as the tails. With this aim, we are intended to

[1] University of Bologna, Italy; brizzi@stat.unibo.it

use a set of symmetric quantiles; we will develop a class of indices, study the corresponding sample distribution and give an example by calculating new indices, referred to a set of geographical and socio-economic variables, considering updated national data of E.U. countries.

2 Ventile-based indices of skewness

Let Y be a quantitative variable, discrete or continuous, let $y_1, y_2, .., y_n$ be the set of data we have to analyse. Denote with $y_{(1)}, y_{(2)}, .., y_{(n)}$ the same data, arranged in non-decreasing order, and let $C_{(k)}$ be the k-th centile of the same data. We could consider every set of quantiles, even the whole set of 99 centiles, but it would belogically weak to calculate such a number of statistics on sample data, especially if the set is not so large. On the other side, focusing our analysis on a reduced set quartiles or deciles) would surely lead us to throw away too much of sample information; we have to find a compromise between simplicity and precision: therefore, we will propose here a ventile-based approach.

From the arranged data $y_{(1)}, y_{(2)}, .., y_{(n)}$, we can easily determine nineteen *sample ventiles*, which correspond to the centiles $C_{(5)}, C_{(10)},\dots$, up to $C_{(95)}$. We will denote the j-th ventile by $V_{(j)}$, following the usual convention to put:

$$V(j) = y(h) \ , \text{ if } \frac{h-1}{n} < \frac{j}{20} < \frac{h}{n} \qquad (2.1a)$$

$$V(j) = \frac{y_{(h)} + y_{(h+1)}}{2} \ , \text{ if } \frac{h}{n} = \frac{j}{20} \ . \qquad (2.1b)$$

As a simple example, we have, for a sample size $n = 25$. The 19 ventiles are the following order statistics:

$y_{(2)}$, $y_{(3)}$, $y_{(4)}$, $[y_{(5)}+y_{(6)}]/2$, $y_{(7)}$, $y_{(8)}$, $y_{(9)}$, $[y_{(10)}+y_{(11)}]/2$, $y_{(12)}$, $y_{(13)}$ *(median)*, $y_{(14)}$, $[y_{(15)}+y_{(16)}]/2$, $y_{(17)}$, $y_{(18)}$, $y_{(19)}$, $[y_{(20)}+y_{(21)}]/2$, $y_{(22)}$, $y_{(23)}$, $y_{(24)}$.

Now, if we take the average of the 19 ventiles, we derive the *Ventile Average (VA)*, a robust estimator of the population mean, belonging to the class of L-statistics (see Hampel *et al.*, 1986). Analogously, we can calculate, directly on ventiles, some indices of dispersion, such as ventile standard deviation (VSD) and ventile absolute deviation about the median (VAD), respectively given by:

$$\text{VSD} = \sqrt{\frac{\sum_{i=1}^{19} (V_{(i)} - VA)^2}{19}} \qquad (2.2)$$

$$\text{VAD}=\frac{\sum_{i=1}^{19}|V_{(i)}-V_{(10)}|}{19} \tag{2.3}$$

We will use these ventile-based statistics in the standardizing procedure, described later. Following the same approach proposed in Brizzi (2000), we can arrange the ventiles in symmetric couples, considering the median apart and take their midvalues:

$$M_{(0)}=V_{(10)},\quad M_{(1)}=\frac{V_{(9)}+V_{(11)}}{2},\ M_{(2)}=\frac{V_{(8)}+V_{(10)}}{2},\ \ldots,\ M_{(9)}=\frac{V_{(1)}+V_{(19)}}{2} \tag{2.4}$$

Following Tukey (1977), we will call these values “midsummaries” If the sample is perfectly symmetric, the midsummaries are all equal. Otherwise, if the data are positively (negatively) skewed, the midsummaries would have an increasing (decreasing) trend. We can then consider the slope of a least-squares straight line interpolating the midsummaries as an index of skewness. Since the values defined in (2.4) depend on the level of magnitude (or unit of measurement) of the data, it is useful to standardize them in order to allow a wider comparison. We suggest two distinct criteria of standardization, based on VSD and VAD, respectively:

$$u_{(i)}=\frac{M_{(i)}-VA}{VSD},\quad i = 0, 1, 2, \ldots,9 \tag{2.5}$$

$$w_{(i)}=\frac{M_{(i)}-M_{(0)}}{VAD},\quad i = 0, 1, 2, \ldots,9 \tag{2.6}$$

If we consider couples of values $(t_{(i)}, u_{(i)})$, where $t_{(i)} = i/10$, and plot them on the plane, we can plot a graphical representation of the skewness of our dataset. Moreover, if we interpolate these points with a straight line, using the standard least squares method, the slope may be a suitable index of skewness; we call it *Ventile Coefficient of Skewness*, defined by:

$$VCS=\frac{Cov(t_i,u_i)}{Var(t_i)} \tag{2.7}$$

Considering that t(i) values are not random at all, we can rewrite the VCS as a linear combination of u(i) values, and precisely:

$$VCS=\frac{18}{33}u_{(9)}+\frac{14}{33}u_{(8)}+\frac{10}{33}u_{(7)}+\frac{6}{33}u_{(6)}+\frac{2}{33}u_{(5)}-\frac{2}{33}u_{(4)}-\ldots-\frac{18}{33}u_{(0)} \tag{2.8}$$

If we do the same with the points $(t_{(i)}, w_{(i)})$, and take the slope of the standard least squares interpolating line, we can define another index of skewness, called here the *Ventile Index of Skewness* (VIS). The formal expression is:

$$VIS = \frac{Cov(t_i, w_i)}{Var(t_i)} \tag{2.9}$$

As well as VCS, also the index VIS may be rewritten as a linear combination of standardized midsummaries:

$$VIS = \frac{18}{33}w_{(9)} + \frac{14}{33}w_{(8)} + \frac{10}{33}w_{(7)} + \frac{6}{33}w_{(6)} + \frac{2}{33}w_{(5)} - \frac{2}{33}w_{(4)} - \ldots - \frac{18}{33}w_{(0)} \tag{2.10}$$

The indices VCS and VIS may be applied directly on theoretical distributions, since their definition is univocal; this was not possible when using letter values, because the definition of the indices depended on the size n. If we apply the new indices to a “classic” positively skewed model, such as negative exponential, we may determine the level of skewness of the distribution itself, thus fixing a reference value, which helps us for an easier interpretation of the indices proposed. The standardizing procedures (2.5) and (2.6) make the indices invariant by linear transformation, and we have then “unique” exponential values of VCS and VIS, regardless of the exponential parameter λ. These “typically exponential” values of the indices are: *VCS= 1.016, VIS= 1.327*. Being the value of VCS very near to one under an exponential distribution, the same index becomes easier to interpret: a value of 0.5, say, indicates almost a half of the skewness corresponding to an exponential model. Moreover, due to the use of ventiles as source of the data information, the statistics VCS and VIS can also be applied to heavy-tailed models like Cauchy or Pareto. Being the Cauchy distribution symmetrical, both the indices are equal to zero; for what concerns Pareto distribution, we have represented some values in Table 1:

Table 1: Ventile-based indices of skewness under a theoretical Pareto model (κ=1, α variable).

α	Nr. of finite moments	VCS	VIS
1	0	1.557	2.963
1.5	1	1.469	2.414
2	1	1.391	2.135
3	2	1.288	1.859
4	3	1.228	1.723
5	4	1.189	1.642

3 Application: The series of prime numbers

We have applied the indices of skewness above defined to a particular "natural" set of values, taken from arithmetics: the set of prime numbers less than *N*, and studied the behavior of skewness as *N* increases.

We have then tried to compare the "classic" moment-based index of skewness (Pearson's γ) with the ventile-based indices, on the set of prime numbers less than N, for some values of N from 100 to 50,000. We have reported, in Table 2, for each limit value *N*, some interesting statistical features: the number *N** of prime numbers in the set, the coverage fraction of prime numbers (*N*/N*), the values taken by the indices γ, VCS and VIS and the ratio VIS/VCS.

Table 2: Moment- and Ventile-based indices of skewness applied to prime numbers.

N	*N**	*N* / N*	*γ*	*VCS*	*VIS*	*VIS/VCS*
100	25	0.250	0.230	0.206	0.238	1.155
300	62	0.207	0.167	0.176	0.201	1.142
500	95	0.190	0.162	0.169	0.195	1.154
1000	168	0.168	0.155	0.171	0.197	1.152
3000	430	0.143	0.128	0.159	0.184	1.157
5000	669	0.134	0.124	0.139	0.161	1.158
10000	1229	0.123	0.111	0.127	0.146	1.144
25000	2762	0.110	0.100	0.115	0.132	1.148
50000	5133	0.103	0.095	0.108	0.125	1.157

Looking at the table, we notice that there is an evident decreasing trend of skewness, with some little oscillation (the series of prime numbers, as usual, has often "weak" regularities). The tendency is almost perfectly similar by considering all the indices shown; sometimes may happen, for small changes, that a decrease of γ corresponds to an increase of ventile-based indices and vice versa (look the values for *N=500, N=1000*). On the other side, the indices VCS and VIS do always move in the same direction, and their ratio results to be approximately constant (about 1.15). It may be also interesting to observe that the values of VCS and VIS are not far from corresponding γ values.

4 Sample distribution and inference

The indices VCS and VIS may be also used within a test of hypothesis regarding population symmetry; if we want to check their performance as test statistics, we need to know – or to estimate – the sample distribution of the above mentioned indices. The sample distribution of VCS and VIS has been studied by a Monte Carlo simulation., performed with “GAUSS” statistical package, under some typical hypotheses on population distribution, corresponding to different levels of skewness.

If we have to deal with unimodally distributed data, the indices of skewness may be used as “quick test statistics” for checking normality. Therefore, we have simulated the sample distribution under the hypothesis of normality; being the indices independent by linear transformation, we considered a standard normal distribution. We have simulated then, for each sample size considered (ranging from *n=15* to *n=75*), 100,000 samples taken form a standard normal population, computing the values of VCS and VIS. We have represented, respectively in Table 3 (VCS) and Table 4 (VIS), the main features of the sample distribution of the ventile-based indices:

Table 3: Sample distribution of VCS under the hypothesis of standard normality.

n	Average	St.Dev.	Centiles: 1.st	2.nd	5.th	95.th	98.th	99.th
15	0.0003	0.4483	-1.003	-0.902	-0.740	0.737	0.899	1.004
18	0.0011	0.4243	-0.952	-0.851	-0.696	0.702	0.859	0.956
25	-0.0003	0.3687	-0.842	-0.750	-0.612	0.606	0.748	0.834
30	0.0002	0.3343	-0.763	-0.679	-0.550	0.549	0.676	0.760
35	-0.0016	0.3087	-0.712	-0.629	-0.509	0.508	0.628	0.704
45	0.0008	0.2794	-0.635	-0.565	-0.461	0.462	0.572	0.640
60	-0.0001	0.2388	-0.548	-0.488	-0.393	0.392	0.485	0.548
75	-0.0012	0.2168	-0.502	-0.446	-0.358	0.356	0.441	0.497

Table 4: Sample distribution of VIS under the hypothesis of standard normality.

n	Average	St.Dev.	Centiles: 1.st	2.nd	5.th	95.th	98.th	99.th
15	0.0015	0,5827	-1.358	-1.196	-0,958	0,955	1,197	1,363
18	0.0015	0.5618	-1.303	-1.150	-0,921	0,929	1,160	1,306
25	-0.0003	0.4614	-1.082	-0.950	-0,763	0,758	0,947	1,074
30	0.0002	0.4177	-0.969	-0.858	-0,686	0,683	0,855	0,969
35	-0.0020	0.3865	-0.904	-0.795	-0,637	0,633	0,791	0,895
45	0.0011	0.3451	-0.796	-0.702	-0,567	0,570	0,710	0,800
60	-0.0002	0.2943	-0.682	-0.605	-0,484	0,482	0,600	0,683
75	-0.0015	0.2670	-0.624	-0.553	-0,440	0,438	0,546	0,616

The simulated sample distributions of VCS and VIS, under a Gaussian model, are approximately symmetric about zero, and the standard deviation is almost linearly proportional to $\sqrt{n}$. Since the inequality **VAD < VSD** holds from well known minimum properties, it is not surprising that VIS (whose standardization is based on VAD) has a larger standard deviation, and tail centiles more distant to zero, than VCS.

We have then calculated the power of VCS and VIS, as test statistics, against a slightly (positively) skewed alternative (Rayleigh distribution), and against a strongly skewed one (negative exponential distribution): for each sample size we have simulated 100,000 samples from a Rayleigh (and then Exponential) distribution and calculated the indices of skewness, checking how many samples did overtake the tail centiles under normality. We have doner the same with Pearson index (γ) as a test statistic, comparing the "old" index with the "new" ones. In Table 5 we have reported the main results.

Table 5: Power of the indices g, VCS and VIS under Rayleigh and Exponential model.

n	Signif. Level	Rayleigh Gamma	Rayleigh VCS	Rayleigh VIS	Exponential Gamma	Exponential VCS	Exponential VIS
15	**0.05**	18.45%	18.32%	**18.84%**	67.51%	**74.05%**	73.25%
	0.01	5.33%	5.17%	**5.71%**	40.25%	**50.31%**	50.06%
18	**0.05**	21.57%	**21.70%**	21.46%	76.67%	**82.57%**	81.71%
	0.01	6.57%	6.45%	**6.64%**	49.80%	**62.62%**	62.19%
25	**0.05**	**28.51%**	21.57%	21.16%	**89.24%**	84.04%	83.22%
	0.01	**9.08%**	6.76%	6.48%	**67.74%**	65.14%	63.72%
30	**0.05**	**33.13%**	26.11%	25.95%	**94.11%**	91.26%	91.78%
	0.01	**11.96%**	9.00%	8.92%	**78.67%**	77.68%	76.55%
35	**0.05**	**37.91%**	29.94%	30.08%	**96.86%**	95.26%	94.86%
	0.01	**13.61%**	10.79%	10.79%	84.88%	**86.01%**	85.11%
45	**0.05**	**46.26%**	32.13%	31.63%	**99.17%**	97.07%	96.82%
	0.01	**19.94%**	12.04%	12.22%	**94.54%**	90.30%	89.60%
60	**0.05**	**58.74%**	42.19%	41.73%	**99.91%**	99.51%	99.44%
	0.01	**27.96%**	18.74%	18.26%	**98.81%**	97.73%	97.39%
75	**0.05**	**69.16%**	48.70%	48.05%	**99.93%**	99.62%	99.56%
	0.01	**37.26%**	23.56%	23.13%	**99.79%**	99.27%	99.18%

In Table 3, we have evidenced in **bold** the maximum power resulting for every combination of alternative distribution, sample size and significance level.

Looking at Table 5, we notice that the new indices (VCS and VIS) are more powerful than γ just for small values of n, whereas the "classic" index γ performas much better for larger values. If we compare the ventile-based indices by means of power, the performances are very similar. For an exponential alternative, VCS is always more powerful than VIS, but the difference is not relevant. In order to increase the power, we propose in the next chapter the "extended" ventile-based indices.

5 Extended indices of skewness

The indices VCS and VIS are robust, because they do not consider at all what happens in the tails; for instance, if the sample size is 75, three data from each tail are "dumb", as they do not have any influence on the indices value. On the other side, this trimming procedure reduces the power of the indices as test statistics. If we want to give back some meaning to the tail values, and to increase the power of the related test of skewness, we can define an *extended index of skewness* corresponding to each ventile-based index, by adding a further midsummary as the extremes midvalue: $M_{(10)} = \frac{y_{(1)} + y_{(n)}}{2}$. This "new" midsummary may be standardized by (2.5) or (2.6), thus extending the series of points representing the skewness. Since this last point covers all the sample, it is quite natural to put the corresponding abscissa $t_{(10)} = 1$.

Table 6: Sample distribution of ECS and EIS under the hypothesis of normality.

n	Average	St.Dev.	Centiles: 1.st	2.nd	5.th	95.th	98.th	99.th
ECS								
15	**-0.0014**	**0.4142**	**-0.927**	**-0.833**	**-0.686**	**0.678**	**0.828**	**0.922**
18	**-0.0029**	**0.3897**	**-0.876**	**-0.789**	**-0.648**	**0.639**	**0.783**	**0.872**
25	**0.0008**	**0.3240**	**-0.737**	**-0.656**	**-0.534**	**0.534**	**0.658**	**0.733**
30	**0.0005**	**0.2978**	**-0.680**	**-0.605**	**-0.490**	**0.491**	**0.605**	**0.679**
35	**0.0012**	**0.2790**	**-0.635**	**-0.563**	**-0.458**	**0.461**	**0.570**	**0.643**
45	**0.0001**	**0.2466**	**-0.562**	**-0.501**	**-0.407**	**0.407**	**0.504**	**0.568**
60	**0.0011**	**0.2179**	**-0.500**	**-0.443**	**-0.357**	**0.361**	**0.446**	**0.501**
75	**0.0007**	**0.1991**	**-0.458**	**-0.407**	**-0.328**	**0.327**	**0.407**	**0.458**
EIS								
15	**-0.0014**	**0.5305**	**-1.244**	**-1.090**	**-0.875**	**0.866**	**1.087**	**1.234**
18	**-0.0038**	**0.5080**	**-1.178**	**-1.047**	**-0.842**	**0.830**	**1.039**	**1.175**
25	**0.0010**	**0.4109**	**-0.957**	**-0.842**	**-0.677**	**0.676**	**0.847**	**0.958**
30	**0.0006**	**0.3776**	**-0.878**	**-0.774**	**-0.621**	**0.622**	**0.775**	**0.878**
35	**0.0017**	**0.3543**	**-0.821**	**-0.723**	**-0.580**	**0.584**	**0.731**	**0.833**
45	**0.0002**	**0.3136**	**-0.725**	**-0.643**	**-0.516**	**0.518**	**0.646**	**0.733**
60	**0.0015**	**0.2784**	**-0.646**	**-0.570**	**-0.455**	**0.461**	**0.574**	**0.649**
75	**0.0010**	**0.2554**	**-0.595**	**-0.526**	**-0.419**	**0.420**	**0.526**	**0.595**

The sample distributions above represented may be used for defining a statistical test for checking the null hypothesis of symmetry.

Applying the standardization (2.5) we derive the *extended coefficient of skewness* (ECS), defined as (2.7), just adding a point; ECS may be written, like VCS, as a linear combination of $u_{(i)}$*'s:*

$$\text{ECS}=\frac{Cov(t_i,u_i)}{Var(t_i)}=\frac{5}{11}u_{(10)}+\frac{4}{11}u_{(9)}+\frac{3}{11}u_{(8)}+\frac{2}{11}u_{(7)}+\frac{1}{11}u_{(6)}-\frac{1}{11}u_{(4)}-...-\frac{5}{11}u_{(0)} \quad (5.1)$$

On the other side, applying the standardization (2.6) we derive the *extended index of skewness* (EIS), defined as (2.9). The EIS may be expressed as:

$$\text{EIS}=\frac{Cov(t_i,w_i)}{Var(w_i)}=\frac{5}{11}w_{(10)}+\frac{4}{11}w_{(9)}+\frac{3}{11}w_{(8)}+\frac{2}{11}w_{(7)}+\frac{1}{11}w_{(6)}-\frac{1}{11}w_{(4)}-...-\frac{5}{11}w_{(0)} \quad (5.2)$$

Once defined these extended indices, we have performed again a simulation, in order to study the sample distribution of VCS and VIS:

Looking at Table 7, we can observe that the extended indices (*ECS*, *EIS*) are more powerful than γ, for every sample size considered and for both the alternatives proposed. The difference seems to be more relevant when considering a reduced significance level *(α=0.01)*. When considering the exponential alternative and a large sample size, the indices are almost equally powerful, since in such conditions the power is very near to one.

Table 7: Power of the indices ECS and EIS under Rayleigh and Exponential model: power percentage and comparison with **γ** .

n	Level of α	Rayleigh ECS	Rayleigh EIS	Rayleigh ECS (γ = 100)	Rayleigh EIS (γ = 100)	Exponential ECS	Exponential EIS	Exponential ECS (γ = 100)	Exponential EIS (γ = 100)
15	**0.05**	20.71%	20.86%	112.25	113.06	78.84%	77.23%	116.78	114.39
	0.01	5.93%	6.16%	111.26	115.57	57.63%	55.26%	143.16	137.27
18	**0.05**	24.58%	24.25%	113.95	112.42	86.60%	85.38%	112.96	111.36
	0.01	7.73%	7.82%	117.66	119.03	69.52%	67.43%	139.60	135.41
25	**0.05**	30.49%	30.26%	106.94	106.14	94.37%	93.69%	105.75	104.99
	0.01	10.96%	10.94%	120.70	120.48	84.28%	82.40%	124.41	121.64
30	**0.05**	36.89%	36.00%	111.35	108.66	97.54%	97.09%	103.65	103.17
	0.01	14.51%	14.28%	121.32	119.40	91.89%	90.61%	116.80	115.17
35	**0.05**	42.27%	41.84%	111.50	110.37	98.99%	98.73%	102.20	101.94
	0.01	17.43%	17.36%	128.07	127.55	95.74%	94.73%	112.79	111.60
45	**0.05**	52.11%	51.09%	112.65	110.44	99.78%	99.71%	100.62	100.55
	0.01	24.98%	24.52%	125.28	122.97	98.82%	98.46%	104.52	104.15
60	**0.05**	65.39%	63.82%	111.32	108.65	99.99%	99.98%	100.08	100.07
	0.01	37.83%	36.09%	135.30	129.08	99.88%	99.82%	101.08	101.02
75	**0.05**	75.28%	73.41%	108.85	106.15	99.999%	99.998%	100.01	100.01
	0.01	48.44%	45.79%	130.01	122.89	99.988%	99.980%	100.20	100.19

6 Application to national data of E.U. countries

Finally, this ventile-based methodology has then been applied to a dataset of national data referred to the 27 countries of E.U. We have chosen a set of eight geographical and socio-economic variables for this application. The variables, labaled form X_1 to X_8, are: area (in squared kms), population (thousands of resident people), income *per capita*, life expectation at birth (years), unemployment rate (in %), diffusion of Personal computers and mobile phones. Finally, we considered also the value of HDI (Human Development Index), a recently-defined index trying to give a normalised measure to human welfare, used since 1990 by the United Nation Development Programme. According to last evaluations, the highest HDI value in the world is 0.963 (Norway), while the lowest is 0.298 (Sierra Leone). In Table 8 we reported the ventile-based statistics and Pearson's index of skewness (γ), in order to make some comparisons.

Table 8: Ventile-based statistics and Pearson's γ for national data of E.U. countries.

Variable		*VA*	*VSD*	*VCS*	*VIS*	*ECS*	*EIS*	γ
Area (sq.kms)	X_1	**155188.7**	**150072.1**	**1.324**	**1.798**	**1.358**	**1.845**	**1.049**
Population (.000)	X_2	**16279.6**	**19004.5**	**1.469**	**2.267**	**1.712**	**2.643**	**1.505**
Income per cap. (EUR)	X_3	**18527.5**	**10726.5**	**0.302**	**0.338**	**-0.934**	**-1.072**	**0.846**
Life expectation	X_4	**77.00**	**2.72**	**-1.118**	**-1.465**	**-1.018**	**-1.334**	**-0.653**
Unemployment Rate (%)	X_5	**7.94**	**2.59**	**0.410**	**0.582**	**0.719**	**0.804**	**1.442**
Pers.Computer (x1000 people)	X_6	**362.95**	**184.80**	**0.310**	**0.359**	**0.759**	**1.078**	**0.275**
Mobile phones (x1000 people)	X_7	**962.05**	**135.91**	**0.407**	**0.520**	**0.354**	**0.411**	**0.532**
H.D.I. (‰)	X_8	**900.58**	**43.07**	**-0.814**	**-0.934**	**0.428**	**0.546**	**-0.679**

Source of data: "Calendario Atlante 2007", Istituto Geografico De Agostini, Novara.

Looking at Table 8, we can point out many important things. First of all, we can use a complete set of ventile statistics (average, standard deviation, skewness) as a brief picture of the behaviour of EU countries with respect to the variables considered here. Focusing our attention on skewness, we can easily notice that all the indices considered are concordant (positive or negative). Moreover, we can make three kinds of comparison between indices:

a) VCS/VIS against Gamma. The most relevant differences are registered for X_3, X_4, X_5. For two of them (X_3 and X_5) γ value is markedly higher; this fact can be explained with the presence of a small number of outliers and the robustness of VCS/VIS with respect to them. For X_4 , γ value markedly lower, and this may be explained (although less clearly) with the low variability of X_4 itself.

b) VCS against VIS. The latter index has always a higher value, due to the different kind of normalisation (VAD is always lower than VSD). For some variable the difference is very relevant, especially for X_2 , which is the variable with the highest level of variability (the only one having VSD > VA) and the highest level of skewness, with respect to all indices.

c) ECS/EIS against VCS/VIS. The values of extended indices are sensibly different to corresponding non-extended ones when considering variables X_3 and X_5. Once again, this is likely due to the presence of outliers (Luxembourg for income, Poland and Slovakia for Unemployment rate), whose effect is reduced (or totally eliminated) by robust indices VCS/VIS, while is kept by extended indices, including the extreme midsummary $M_{(10)}$. However, as stated before, the extended indices are to be considered more as a test statistic than an exploratory tool.

7 Final comments

The indices VCS and VIS, introduced and developed here, are simple, robust and easy to interpret statistics, suitable for checking the skewness of a set of data, as well as the extended indices ECS and EIS are a powerful tool for making inference about symmetry. The indices, as pointed out in this paper, may be used even for evaluating data coming from heavy tailed distributions. This method for defining indices, developed here for ventiles, could be easily generalised to other sets of symmetric quantiles (deciles, centiles or whatever else). We have considered, in this study, that ventiles may be a possible compromise between simplicity and precision; nonetheless, any other choice is undoubtedly worth of attention. It would be interesting, in a further research, to make a comparison between the performances of indices resulting from each choice of quantiles, and to compare all them with γ and other existing indices of skewness.

References

[1] Antille, A., Kersting, G., and Zucchini, W. (1982): Testing symmetry. *Journal of the American Statistical Association*, **77**, 639-646.

[2] Arnold, B.C. and Groeneveld, R.A. (1995): Measuring skewness with respect to the mode. *The American Statistician*, **49**, 34-38.

[3] Balanda, K.P. and Mac Gillivray, H.L. (1988): Kurtosis: a critical review. *The American Statistician*, **42**, 111-119.

[4] Brizzi, M. (2000): Detecting skewness and kurtosis by letter values: a new proposal. *Statistica*, LX, 243-258.

[5] Brizzi, M. (2002): Testing symmetry by an easy-to-calculate statistic based on letter values. *Metodoloski Zvezki*, **17**, 63-74**.**

[6] Groeneveld, R.A. and Meeden, G. (1984): Measuring skewness and kurtosis. *The Statistician*, **33**, 391-399.

[7] Groeneveld R.A. (1998): A class of quantile measures for kurtosis. *The American Statistician,* **52**, 325-329.

[8] Hampel, F.R., Ronchetti, E.M., Rousseeuw, P.J., and Stahel W.A. (1986): *Robust Statistics. The Approach Based on Influence Function*. New York: John Wiley & Sons.

[9] Hoaglin, D.C., Mosteller, F., and Tukey, J.W. (1985): *Exploring Data Tables, Trends and Shapes*. New York: John Wiley & Sons.,Chapter 10 by D.C. Hoaglin.

[10] Joanes, D.N. and Gill, C.A. (1998): Comparing measures of sample skewness and kurtosis. *The Statistician*, **47**, 183-189.

[11] Kappenman, R.F. (1988): Detection of symmetry or lack of it and applications. *Communications in Statistics. Theory and Methods*, **17**(12), 4163-4177.

[12] Mac Gillivray, H.L. (1986): Skewness and asymmetry: measures and orderings. *Annals of Statistics*, **14**, 994-1011.

[13] Moors, J.J.A. (1988): A quantile alternative for kurtosis. *The Statistician*, **37**, 25-32.

[14] Oja, H. (1981): On location, scale, skewness and kurtosis of univariate distributions. *Scandinavian Journal of Statistics*, **8**, 154-168.

[15] Ruppert, D. (1987): What is Kurtosis? *The American Statistician*, **41**, 1-5.

[16] Tukey, J.W. (1977): *Exploratory Data analysis*. Reading MA: Addison – Wesley.

[17] Wang, J. and Serfling, R. (2005): Nonparametric multivariate kurtosis and tailweight measures. *Journal of Nonparametric Statistics,* **17**, 441-456.