
KSG: Indicator of Economic Concentration

Manuel Meireles*

¹ Master's and Doctoral Departments of UniFaccamp, Campo Limpo Paulista, SP, Brazil, meireles@faccam.br

*Manuel Meireles, Rua: Guatemala, nº 167, Jardim América - CEP 13231-230 Campo Limpo Paulista, SP, Brazil.

ABSTRACT

In the fields of Management and Economics, there are many studies that have made use of the degree of concentration of a market or industry, especially when dealing with subjects such as industrial concentration. However, these indexes do not adequately present the level of significance. This problem is overcome by the proposed KSG indicator based on the Kolmogorov-Smirnov test and whose interpretation of significance is given by Goodman. Hence the name: KSG. The proposed model uses non-parametric techniques to establish the dimension of concentration and defines the level of significance of the value found. This is a quantitative study using parametric statistics (polynomial regression) on data generated through simulation. In each data simulation, for the given value of n companies, the share of Company 1 is made to vary, with the other shares being maintained unchanged. For each simulation, data related to values of "Share of Company 1" were extracted and corresponding indexes: KSG, CR4, CR8 and HHI. The results show that the indicator proposed in this study is fully justified.

Keywords: Indicator of economic concentration; Competitiveness; Imperfect markets; Market concentration; KSG

1. Introduction

This study looks at a proposed indicator of economic concentration designated KSG (derived of Kolmogorov-Smirnov and Goodman) and which has the same potential for application as one of the indicators most currently in use: the Herfindahl-Hirschman index (HHI). The concept of concentration, in accordance with Farina (1993), has to do with not only the number of participants in a given market. More importantly, it has to do with but to the control of a large portion of a given economic activity by a small number of participants in this activity. According to Bem (1991), concentration can be defined as the measure of firms in order of size (this can be measured in terms of sales, number of employees or production capacity), ranging from the largest to the smallest company.

In the fields of Management and Economics, there are many studies that have made use of the degree of concentration of a market or industry, especially when dealing with themes such as industrial concentration, mergers, turnover, market share, monopolies and oligopolies. Those that deserve to be mentioned include the studies of Lerner^[1], Iwata^[2], Braga and Mascolo^[3], Tirole (p. 247)^[4], Scherer and Ross^[5], Tonge and Wootton^[6], Baker and Bresnahan^[7], Cuevas^[8], Tiffin and Dawson^[9], Borenstein *et al.*^[10], Mahajan^[11], Ballas and Fafaliou^[12], Kozyrev and Malyzhenkov^[13] and Al-Jarrah *et al.*^[14]

According to Kon^[15], measuring concentration supplies the empirical elements necessary to evaluate the situation of competition in a market and also serves to make inter-temporal comparisons that enable an examination of the market dynamic in addition to the offer that facilitates strategic decision making.

2. Conceptual Aspects

The most common measures of the concentration of an industry, according to Kon^[15], are: the concentration ratio (Ck); the HHI. There are other indexes for this purpose, as can be seen in Kupfer and Hasenclever^[16], Clarke^[17] and Tabner^[18]. A summary of the main indexes of concentration is shown in Chart 1 and described in the following paragraphs.

Concentration Indicator	Symbol	Author
Herfindahl-Hirschman Index	HHI	Hirschman ^[19] , Herfindahl ^[20]
Simpson's Index	E	Simpson ^[21] , Yule ^[22] , Fisher <i>et al.</i> ^[23] , Williams ^[24]
Hannah and Kay's Index	R	Hannah and Kay ^[25]
Shannon's Entropy Index	J	Shannon ^[26] ; Shannon & Weaver ^[27] ; Nyquist ^[28] ; Hartley ^[29] ; Wiener ^[30]
Gini Coefficient	G	Gini ^[31]
Coefficient of Variation of firm size	Cv	Clarke ^[17] (derived from HHI)
Concentration ratio	Ck	Unknown author, mentioned by Clarke ^[17]
Variance of the logarithms of firm size	V ²	Unknown author, presented by Clarke ^[17]
Lorenz Curve		Lorenz ^[32]
Concentration Curve		Clarke ^[17]

Chart 1. Indicators of Economic Concentration

Source: Developed by the author

Herfindahl-Hirschman Index (HHI). This index is widely used by competition regulators. It was initially attributed to Herfindahl^[20] in a dissertation at Columbia University. However, a similar indicator had been presented by Hirschman^[19], expressed in a different way. The HHI is calculated as the sum of the squares of the market share of the companies that operate in the market in question and varies from 0 to 10,000. A zero value means that there is no company in a given market. The formula for this calculation is:

$$HHI = \sum_{i=1}^n 10000 \left(\frac{w_i}{w} \right)^2 \quad (1)$$

where w_i is the percentage share of the market w of each company. According to HMG^[33], the State Agency on Antimonopoly of the United States Justice Department divides the spectrum of market concentration, measured by the HHI, into three regions, as shown in **Table 1**.

Observed HHI	Interpretation
$1000 \leq HHI$	Unconcentrated market
$1000 < HHI \leq 1800$	Moderately concentrated
$HHI > 1800$	Highly concentrated market

Table 1. Interpretation of HHI value

Source: HMG^[33]

Simpson's Index (E). The Simpson index was initially proposed by Simpson^[21], although Yule^[22] had presented a similar indicator known as G. Simpson combined the ideas of Yule with those of Fisher *et al.*^[23] and Williams^[24] to create the basis of this index. The value varies from 0 to 1, with 1 meaning no concentration. The formula for this calculation is:

$$E = \left(\frac{1}{N} \right) \left(\frac{1}{\sum_{i=1}^N \left(\frac{w_i}{w} \right)^2} \right) \quad (2)$$

where N is the number of industries and w is the weight of each industry.

Hannah and Kay's Index (R). Hannah and Kay's index^[25] is similar to the HHI, especially when $\alpha=2$, but it has the advantage of allowing a choice of which part of the concentration curve to focus on, changing the value of α arbitrarily. The higher the value of this parameter, the greater will be the weight attributed to the firms according to their size. It is calculated as follows:

$$R = \sum_{i=1}^N w_i^\alpha \quad \alpha > 0 \quad (3)$$

When $\alpha=2$, the value of the R index is similar to that of the HHI.

Shannon's Entropy Index (J). Shannon's is also known as the entropy index. It is calculated using the following formula:

$$J = -\sum_{i=1}^N w_i \ln w_i = J = \sum_{i=1}^N w_i \ln \left(\frac{1}{w_i} \right) \quad (4)$$

The best value of J occurs when the share of companies is equal and is given by $J_{ideal} = \ln(N)$. The Shannon index was later discussed by Shannon and Weaver^[27] and later still by Fernholz *et al.*^[34]. In texts regarding diversity, this indicator is known as the Shannon-Wiener Index in deference to Wiener^[30], who arrived independently at a similar index in 1948.

Gini Coefficient (G). The G coefficient was proposed by Gini^[31]. It consists of a number between 0 and 1, where 0 corresponds to complete equality and 1 corresponds to complete inequality. The Gini index is the coefficient expressed in percentage points (equal to the coefficient multiplied by 100). The Gini coefficient is calculated according to Formula 5, which is expressed as the ratio of the areas in the diagram of the Lorenz Curve. If the area between the line of perfect equality and the Lorenz Curve is A, and the area below the Lorenz Curve is B, then the Gini Coefficient is equal to $A/(A+B)$.

$$G = \frac{\sum_{i=1}^N \sum_{j=1}^N |x_i - x_j|}{2N^2 \mu} \quad (5)$$

If $G=0$, there is an equal share of the attribute; if $G=1$ there is maximum concentration; for $0 \leq G \leq 1$ G is greater the higher the concentration.

Coefficient of Variation of Firm Size (Cv). This coefficient is the ratio of the standard deviation of the firm's market share and the median of this share. According to Clarke^[17], as the standard deviation is expressed in the same unit of data observed in the study, it is impossible to compare two or more series of value that are in different measuring units. To overcome these difficulties, the dispersion can be analyzed in relative terms to its median value, using the Pearson coefficient. The variation of firm size coefficient is arrived at using the following formula:

$$Cv = 100 \frac{s}{\bar{w}} = 100 \frac{\sqrt{\frac{(w_i - \bar{w})^2}{n-1}}}{\frac{\sum_{i=1}^n x_i}{n}} \quad (6)$$

Concentration Ratio (Ck). The concentration ratio (Ck) measures the proportion represented by a number k of the largest companies in an industry in relation to the total. It considers the share of the largest firms in the total, i.e., the ratio of the largest k firms in a market with n firms will be defined as:

$$C_k = \frac{\sum_{i=1}^k w_i}{\sum_{i=1}^n w_i} \quad (7)$$

with k = the number of firms in the calculation; n = the total number of companies in the market and w_i = share of the i-th firm in the market. This index is easy to interpret and indicates the market share (concentration of market power) of

the k largest companies that make up the industry. The concentration ratio is set by decreasingly ordering the variable under study (percentage of total sales, for example).

Variance of the Logarithms of Firm Size. This indicator is arrived at using the following equation:

$$V^2 = \frac{1}{N} \sum_{i=1}^N \left[\log \left(\frac{x_i}{\bar{x}_g} \right) \right]^2 \quad (8)$$

where \bar{x} is equal to the geometric median of the size of the firms. This equation is particularly useful, according to Clarke^[17], if the distribution of size of firm follows the normal log distribution. This author claims that in a situation in which the size of the forms approximately follows a normal log distribution and the Lorenz curves do not intersect, V^2 enables an unequivocal measure of concentration.

Lorenz Curve. The Lorenz Curve, or Lorenz concentration curve, developed by Max Lorenz^[32] in 1905, consists of a graph that seeks to illustrate the inequality that exists in the distribution of a given variable. In this diagram, a diagonal line represents the theoretical equitable distribution (t) and another line represents the observed distribution (o), as shown in **Figure 1**. The farther the diagonal is from the line of observed values, the higher the concentration of the variable under study, or the greater will be the inequality between the values observed and the theoretical values. The Lorenz Curve can be complemented by the Gini index, which quantifies the degree of concentration of income.

Concentration Curve. The concentration curve is an absolute measurement of concentration in which the inequality of size of firms is represented by the convexity of a curve while a straight line indicates the theoretical share of firms. Clarke^[17] shows an example, described in **Figure 2**, in which the cumulative percentage of the share of companies is represented by a curved line (a), while the straight line (b) indicates the theoretical cumulative share if all the companies had an equal share. The higher the concentration, the greater is the convexity; when concentration is nil, line (o) grows closer to line (t).

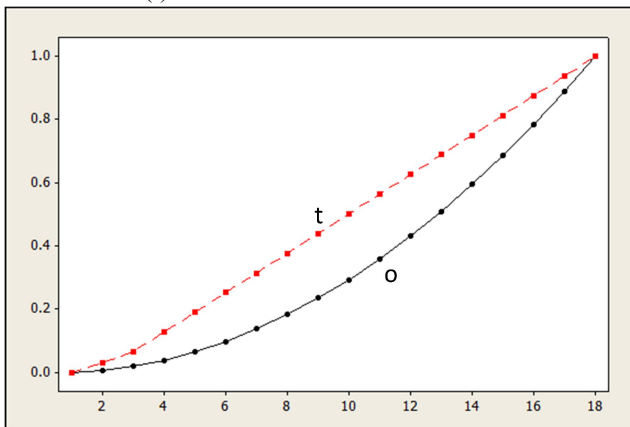


Figure 1; Lorenz Curve.

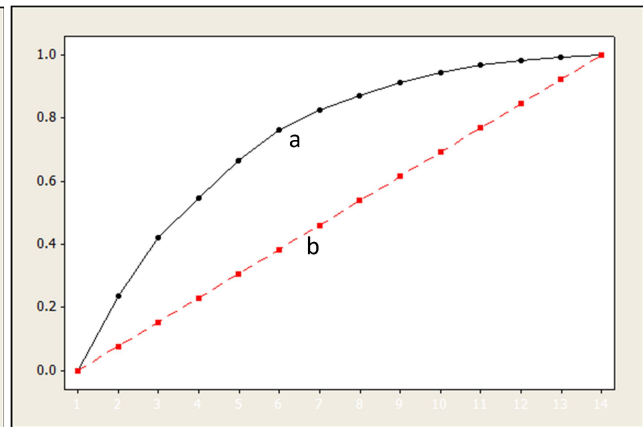


Figure 2; Concentration Curve.

Source: Clarke^[17]

3. Hypotheses

The concentration ratio (CK) indexes and the Herfindahl-Hirschman (HHI) are used a great deal^[15]. However, these indexes do not adequately present the level of significance. This inconvenience is overcome by the proposed KSG indicator based on the KS (Kolmogorov-Smirnov) test and whose interpretation of significance is given by Goodman and Kruskal^[35]. Hence the name: KSG.

It is desirable that a new indicator capable of measuring the concentration of companies in relation to a given industry should be consistent with the most currently used indicator, the HHI. Therefore, the current study tests the hypothesis below, given in an alternative form:

H₁: There is no significant difference between the results of the KSG and the HHI for a given n size market.

The proposal for a concentration indicator that is simply consistent with one of the most applicable would not be justified if there were no gain in information. It is here that the quality of the KSG concentration index can be seen in

relation to the HHI. For this reason, the second proposal is formulated:

H₂: The results shown by the KSG have greater informational content than the results of the HHI.

4. Proposed Model

This study proposes a method for defining the degree of concentration based on the Kolmogorov-Smirnov (KS) test for two samples with determined significance in accordance with Goodman and Kruskal^[35]. It begins on the assumption that the relevant market is defined, i.e., the limits and characteristics of the industry whose concentration is going to be measured have already been defined. It is worth highlighting that a market, according to Rubinfeld^[36], does not have to, indeed cannot, be defined with scientific precision. The definition of market is, at the end of the day, a matter of business reality, of how the market is perceived by those who struggle to profit from it.

The proposed model uses non-parametric techniques to establish the dimension of concentration and define the level of significance of the value found. These non-parametric tests play the same roles as parametric tests. The difference is that the non-parametric data do not require data with normal distribution nor other specific conditions that must be met to apply parametric tests.

The advantages of the non-parametric techniques identified by Baquero (p. 216)^[37] are: (1) the normality of the population and their respective parameters is precluded; (2) they are applied to samples obtained from a different population; (3) the measures can be simple and on a “nominal scale; (4) the samples can be really small, and (5) the mathematical calculations are simple.

The present method of calculation assumes that it is possible to obtain for every economic agent his market share (w_i), setting out on the principle that the number n of inherent agents in the market under study is known. In this case, the theoretical share (T_i) of each agent in conditions of perfect competition is defined as $T_i=1/n$. therefore, with decreasing p_i share values, Chart 2 can be constructed.

A	B	C	D	E	F	G
Economic agent A _i	Value for agent in the market V _i	Observed w _i share	Accumulated share A _{wi}	Theoretical T _i share	Accumulated theoretical A _{ti} share	Difference D _i
A1	V1	w1=V1/VT	Aw1=w1	T	T	D1=Aw1-T
A2	V2	w2=V2/VT	Aw2=w1+w2	T	2T	D2=Aw2-2T
A3	V3	w3=V3/VT	Aw3=w1+w2+w3	T	3T	D3=Aw3-3T
...
A(n-1)	V(n-1)	w(n-1)=V(n-1)/VT	Aw(n-1)=w1+w2+...+w(n-1)	T	(n-1)T	D(n-1)=Aw(n-1)-(n-1)T
An	Vn	wn=Vn/VT	1	T	1	0
Totais	VT	1		1		D max
						KSG=2n(Dmax) ²

Chart 2. Proposed Model

Legend: A: name of economic agent; B: value for agent in the market; the total of column pT is the sum of the p_i values; C: observed w_i share relative to each agent; D: observed accumulated share; E: theoretical T_i share of each agent; F: accumulated theoretical share; G: difference between columns D and F: accumulated observed share and accumulated theoretical share; D Max= largest observed D_i difference. **Source:** Author.

The KS test is very similar to the χ^2 (chi-squared). According to Baquero (p.227)^[37], one of the specific elements of the KS is the accumulation of frequencies obtained from a given sample:

What we are trying to solve is whether the differences obtained in a sample are really significant in terms of a determined distribution. If the samples are taken from the same population, the accumulated frequencies of both samples should be the same. What we intend to do, therefore, is demonstrate whether the differences between the accumulated frequencies of the samples are significantly different from the theoretical distribution.

In accordance with the Kolmogorov-Smirnov test, according to Siegel (p.144)^[38], it is possible to verify whether there is agreement between the two cumulative distributions. The test focuses on the largest of the observed deviations:

$$D = \max |iT - APi| \quad (9)$$

According to Siegel (p.146)^[38] sample distribution D is known (Smirnov^[39]). On the other hand, the probabilities

associated with the occurrence of values as large as an observed D under a null hypothesis are tabulated (with both samples coming from the same distribution).

According to Motulsky (p.18)^[40], the Kolmogorov and Smirnov method originally published by these researchers cannot calculate the p value because this requires that the median and standard deviation of the program be known, which is rarely the case. For this reason, the interpretation of Goodman and Kruskal^[35] is adopted. Siegel (p.149)^[38] affirms that Goodman and Kruskal^[35] showed that

$$\chi^2 = 4D^2 \frac{n_1 n_2}{n_1 + n_2} \quad (10)$$

Has an approximately chi-squared sample distribution with $gl=2$. In other words, we can determine the significance of an observed D value. This author highlights (p.153) that it is interesting to note that the chi-squared approximation can also be used for small samples but, in this case, its application leads to conservative proof. In other words, the error committed when applying the chi-squared approximation to small samples will always fall on the side of safety (Goodman and Kruskal^[35]).

This means that the approximation of Goodman and Kruskal^[35] (Formula 10) is applicable to both small and large samples. In the present proposal, as $n_1=n_2$, the following formula can be written:

$$\chi^2 = 4D^2 \frac{n_1 n_2}{n_1 + n_2} = 4D^2 \frac{n^2}{2n} = 2D^2 n = KSG \quad (11)$$

To interpret Formula 11, the chi-squared table of critical values is used, with $gl=2$ and that presented in Table 2 in accordance with Morettin and Bussab^[41].

Probability p, under Ho, of $\chi^2 \geq$ observed									
gl=2	Non-significant						Significant		
Signif.	0.99	0.98	0.95	0.90	0.50	0.10	0.05	0.01	0.001
χ^2	0.02	0.04	0.1	0.21	1.39	4.6	5.99	9.21	13.82

Table 2. Critical values of χ^2

Source: Morettin and Bussab^[41]

Note that the levels of significance ≥ 0.05 are considered significant for what can be attributed in the observance of values $\chi^2 \geq 5.99$ the existence of significant market concentration. It is recommended that there should be an agreement as to the terminology that can be applied concerning the degree of concentration.

To simplify the expression, the designation of KSG is given to the value of Formula 11, expressing the process of the method: Kolmogorov-Smirnov-Goodman, i.e.:

$$KSG = 2D^2 n \quad (12)$$

Table 3 shows nine characteristics of concentration depending on observed KSG that is the same as the observed χ^2 . This table was defined by the author based on the critical values of χ^2 in Table 2. The level of “significant concentration” was defined for significant equal to or lower than 0.05; and the “free competition” level was defined for significance equal to or greater than 0.95.

Observed KSG	Level of significance	General Level	Specific level
≥ 13.82	≤ 0.001	Significant concentration	Absolute
$\geq 9.21 < 13.82$	0,01		Very high
$\geq 5.99 < 9.21$	0,05		High
$\geq 4.6 < 5.99$	0,10	Imperfect competition	Tending to concentrate
$\geq 1.39 < 4.6$	0,50		Intermediate
$\geq 0.21 < 1.39$	0,90		Tendency to free competition
$\geq 0.10 < 0.21$	0,95	Free competition	High

≥ 0.04	< 0.10	0,98	Very high
< 0.04		≥ 0.99	Absolute

Table 3. Interpretation of the observed KSG value

Source: Author

5. Research Method

This is a quantitative study using parametric statistics (polynomial regression) on data generated through simulation.

In each data simulation, for a given value of n companies, the share of Company 1 is made to vary, with the other shares being maintained unchanged. For each simulation, data related to values of “Share of Company 1” were extracted and corresponding indexes: KSG, CR4, CR8 and HHI. Special attention was paid to the cutoff values, both in the proposed model and the HHI index.

The data, especially the KSG and HHI variables, were subjected to polynomial regression analysis. For this analysis, the Minitab R14 statistical package was used.

6. Results

The results were considered in four stages. Initially, the simulations were considered. In this stage, four markets were simulated, varying in number of companies: 9, 33, 99 and 199. The graphic analysis was then considered, correlating the values obtained for the KSG indicator with the HHI values. In the third stage, statistical analyses were conducted and, finally, the cutoff points of the two systems (KSG and HHI) were investigated.

6.1 Simulations

The 45 simulations proceeded and their results are shown in **Table 4**, using four different n values of firms with a share in the market: 9 firms, 33, 66 and 177. These numbers were chosen at random.

The important columns to this study especially are columns n, KSG and HHI, which are linked to the research hypotheses. For example, in simulation line 5 the market is composed of nine forms (n=9), all with a production of 200 units, except Firm 1, which produces 900 units. The value of the KSG index is 1.115 which, interpreted in accordance with Table 3, indicates a specific market level of “tending towards free competition”. The market concentration in this case is not significant, i.e., the level of significance is higher than 0.05. The HHI calculated for this example was 1808.00. In accordance with Table 1, this indicates moderate concentration. It is worth observing that the HHI index is basically passing from “unconcentrated market” to “moderate concentration”. In the interpretation of the KSG index, the value observed is between “free competition” and “imperfect competition”.

Significance	Critical Values	n	Simu-latio n	Value of Emp. 1	KSG	C4	C8	HHI
≥ 0.99	< 0.04	9	1	200	0.000	0.4444	0.8889	111.11
≥ 0.99	< 0.04	9	2	300	0.039	0.4737	0.8947	1135.73
0.98	≥ 0.04 < 0.10	9	3	360	0.095	0.4898	0.8980	1170.35
0.95	≥ 0.10 < 0.21	9	4	445	0.204	0.5110	0.9022	1238.69
0.90	≥ 0.21 < 1.39	9	5	900	1.115	0.6000	0.9200	1808.00
0.90	≥ 0.21 < 1.39	9	6	1015	1.381	0.1676	0.9235	1974.53
0.50	≥ 1.39 < 4.6	9	7	2570	4.594	0.7602	0.9520	3982.37
0.10	≥ 4.6 < 5.99	9	8	3520	5.980	0.8047	0.9609	4848.63
0.05	≥ 5.99 < 9.21	9	9	7610	9.206	0.8914	0.9783	6865.04
0.01	≥ 9.21 < 13.82	9	10	3640	9.213	0.8933	0.9791	6869.02
≤ 0.001	≥ 13.82	9	11	125000	13.821	0.9921	0.9984	9749.63
≥ 0.99	< 0.04	33	12	200	0.000	0.1212	0.2424	303.03
≥ 0.99	< 0.04	33	13	370	0.039	0.1433	0.2614	309.14

0.98	≥0.04 <0.10	33	14	475	0.099	0.1564	0.2727	318.55
0.95	≥0.10 <0.21	33	15	605	0.207	0.1720	0.2862	335.44
0.90	≥0.21 <1.39	33	16	1360	1.387	0.2526	0.3557	519.72
0.50	≥1.39 <4.6	33	17	2665	4.589	0.3602	0.4484	1020.06
0.10	≥4.6 <5.99	33	18	3170	5.977	0.3939	0.4775	1236.98
0.05	≥5.99 <9.21	33	19	4330	9.194	0.4595	0.5340	1739.63
0.01	≥9.21 <13.82	33	20	4475	9.590	0.4667	0.5402	1801.51
0.01	≥9.21 <13.82	33	21	6095	13.814	0.5358	0.5998	2461.43
≤0.001	≥13.82	33	22	8000	18.209	0.5972	0.6528	3148.15
≥0.99	<0.04	66	23	200	0.000	0.0606	0.1212	151.52
≥0.99	<0.04	66	24	435	0.039	0.0770	0.1366	154.53
0.98	≥0.04 <0.10	66	25	575	0.098	0.8660	0.1455	159.03
0.95	≥0.10 <0.21	66	26	755	0.208	0.0985	0.1567	167.55
0.90	≥0.21 <1.39	66	27	1735	1.389	0.1585	0.2128	258.39
0.50	≥1.39 <4.6	66	28	3285	4.595	0.2386	0.2877	504.95
0.10	≥4.6 <5.99	66	29	3840	5.982	0.2637	0.3112	611.65
0.05	≥5.99 <9.21	66	30	5035	9.202	0.3124	0.3568	859.35
0.01	≥9.21 <13.82	66	31	6655	13.809	0.3691	0.4098	1213.74
≤0.001	≥13.82	66	32	9350	21.459	0.4452	0.4810	1802.17
≥0.99	<0.04	177	33	200	0.000	0.0226	0.0452	56.50
≥0.99	<0.04	177	34	580	0.039	0.0330	0.0553	57.62
0.98	≥0.04 <0.10	177	35	805	0.099	0.0390	0.0612	59.30
0.95	≥0.10 <0.21	177	36	1085	0.208	0.0464	0.0685	62.41
0.90	≥0.21 <1.39	177	37	2580	1.389	0.0842	0.1053	95.96
0.50	≥1.39 <4.6	177	38	4780	4.593	0.1346	0.1546	186.99
0.10	≥4.6 <5.99	177	39	5520	5.974	0.1503	0.1699	226.22
0.05	≥5.99 <9.21	177	40	7050	9.200	0.1811	0.2000	317.87
0.01	≥9.21 <13.82	177	41	8975	13.811	0.2168	0.2349	448.85
≤0.001	≥13.82	177	42	16000	33.332	0.3242	0.3398	1003.42
≤0.001	≥13.82	177	43	25800	61.646	0.4328	0.4459	1807.79

Table 4. Results of 43 simulations

Legend: Significance = level of significant referring to the KSG index; the level is considered significant when it is 0.05 or less. Critical values = interval of critical values for χ^2 , in accordance with Table 2, which is the same as the KSG; n= number of firms considered to be participating in the market; Simulation: simulated number; Value of company 1= In the simulation Company 1 has the production of indicated units and each of the other companies has a production of 200 units; KSG = proposed KSG index calculated in accordance with Formula 11; C4 = concentration index CK with k=4 calculated in accordance with Formula 6; C8 = likewise with k=8; HHI = Herfindahl-Hirschman Index calculated in accordance with Formula 1. **Source:** Author.

Table 5 shows in detail Simulation 1, with n=9, and it is a case that expresses free competition. In this simulation, all the companies have the same value and the KSG value is null and interpreted as “absolute free competition” in accordance with Table 3. In the case of HHI, the value 1111.91 is interpreted as “unconcentrated”. The concentration indexes (Ck) with k=4 and k=8 measure the proportion represented by the number k of larger companies in an industry in relation to the total: C4=44.44% and C8=88.89% indicate an equal share.

Table 6 shows Simulation 20 in detail, with 33 firms and a situation where a high concentration begins to exist in accordance with the HHI, which is 1801.51. By the KSG, the value 9.59 indicates “very high significant concentration”

at the significance level of 0.01 as shown in Table 3. The concentration indexes (Ck) with k=4 and k=8 measure the proportion represented by the number k of the largest companies in an industry in relation to the total: C4=46.67% and C8=54.02%. In the case of an equal share, each firm would have 1/33 share (3.003%) and the first four should have a share of 12.12%. Likewise, the first each companies, with an equal share, should have 8/33 (24.24%).

Simulation 1: (n=9)										
Firm	Production	w _i	Aw _i	T _i	At _i	D	C4	C8	HHI	
Firm 1	200	0.1111	0.1111	0.1111	0.1111	0.0000	0.1111	0.1111	123.46	
Firm 2	200	0.1111	0.2222	0.1111	0.2222	0.0000	0.1111	0.1111	123.46	
Firm 3	200	0.1111	0.3333	0.1111	0.3333	0.0000	0.1111	0.1111	123.46	
Firm 4	200	0.1111	0.4444	0.1111	0.4444	0.0000	0.1111	0.1111	123.46	
Firm 5	200	0.1111	0.5556	0.1111	0.5556	0.0000		0.1111	123.46	
Firm 6	200	0.1111	0.6667	0.1111	0.6667	0.0000		0.1111	123.46	
Firm 7	200	0.1111	0.7778	0.1111	0.7778	0.0000		0.1111	123.46	
Firm 8	200	0.1111	0.8889	0.1111	0.8889	0.0000		0.1111	123.46	
Firm 9	200	0.1111	1.0000	0.1111	1.0000	0.0000			123.46	
Total	1800					Maximum D=	0.0000	0.4444	0.8889	1111.11
						KSG	0.000			

Table 5. Simulation 1: values for cases of free competition with an equal share for all firms.

Legend: Firm= “Name of Firm”; Production: quantity of units produced by the firm, for instance; w_i= observed share of each firm in the market; Aw_i= accumulated observed share; T_i= theoretical share of each firm in the case of equal production; At_i= accumulated theoretical production; D= difference between accumulated values: Aw_i-T_i; C4= concentration index CK with k=4 calculated in accordance with Formula 6; C8= likewise with k=8; HHI= Herfindahl-Hirschman index calculated in accordance with Formula 1. Maximum D: maximum observed D value in n cases; KSG= proposed KSG index calculated in accordance with Formula 11: Source: Author.

Simulation 20: (n=33)										
Firm	Prod-uction	w _i	Aw _i	T _i	At _i	D	C4	C8	HHI	
Firm 1	4475	0.4115	0.4115	0.0303	0.0303	0.3812	0.4115	0.4115	1693.28	
Firm 2	200	0.0184	0.4299	0.0303	0.0606	0.3693	0.0184	0.0184	3.38	
Firm 3	200	0.0184	0.4483	0.0303	0.0909	0.3574	0.0184	0.0184	3.38	
Firm 4	200	0.0184	0.4667	0.0303	0.1212	0.3455	0.0184	0.0184	3.38	
Firm 5	200	0.0184	0.4851	0.0303	0.1515	0.3335		0.0184	3.38	
Firm 6	200	0.0184	0.5034	0.0303	0.1818	0.3216		0.0184	3.38	
Firm 7	200	0.0184	0.5218	0.0303	0.2121	0.3097		0.0184	3.38	
Firm 8	200	0.0184	0.5402	0.0303	0.2424	0.2978		0.0184	3.38	
Firm 9	200	0.0184	0.5586	0.0303	0.2727	0.2859			3.38	
Firm 10	200	0.0184	0.5770	0.0303	0.3030	0.2740			3.38	
Firm 32	200	0.0184	0.9816	0.0303	0.9697	0.0119			3.38	
Firm 33	200	0.0184	1.0000	0.0303	1.0000	0.0000			3.38	
Total	10875					Maximum D=	0.3812	0.4667	0.5402	1801.51
						KSG	9.590			

Table 6. Simulation 20, with 33 firms.

Legend: Firm= “Name of Firm”; Production: quantity of units produced by the firm, for instance; w_i= observed share of each firm in the market; Aw_i= accumulated observed share; T_i= theoretical share of each firm in the case of equal production; At_i= accumulated

theoretical production; D= difference between accumulated values: $Aw_i - T_i$; C4= concentration index CK with $k=4$ calculated in accordance with Formula 6; C8= likewise with $k=8$; HHI= Herfindahl-Hirschman index calculated in accordance with Formula 1. Maximum D: maximum observed D value in n cases; KSG= proposed KSG index calculated in accordance with Formula 11: **Source:** Author.

Table 7 shows in detail Simulation 37, with 77 firms and shows a situation in which the market is not concentrated as the HHI is 95.96. By the KSG index, the value 1.389 indicates that the market has a “tendency towards free competition” according to Table 3, at a significance level of 0.90. The concentration observed by the concentration indexes (Ck) with $k=4$ and $k=8$ is given by: C4=8.42% for a theoretical value of $4/177$ (2.26%) and C8=10.532% for a theoretical value of $8/177$ (4.52%).

Simulation 37: (n=177)									
Firm	Production	w_i	Aw_i	T_i	At_i	D	C4	C8	HHI
Firm 1	2580	0.0683	0.0683	0.0056	0.0056	0.0626	0.0683	0.0683	46.64
Firm 2	200	0.0053	0.0736	0.0056	0.0113	0.0623	0.0053	0.0053	0.28
Firm 3	200	0.0053	0.0789	0.0056	0.0169	0.0619	0.0053	0.0053	0.28
Firm 4	200	0.0053	0.0842	0.0056	0.0226	0.0616	0.0053	0.0053	0.28
Firm 5	200	0.0053	0.0895	0.0056	0.0282	0.0612		0.0053	0.28
Firm 6	200	0.0053	0.0948	0.0056	0.0339	0.0609		0.0053	0.28
Firm 7	200	0.0053	0.1001	0.0056	0.0395	0.0605		0.0053	0.28
Firm 8	200	0.0053	0.1053	0.0056	0.0452	0.0601		0.0053	0.28
Firm 9	200	0.0053	0.1106	0.0056	0.0508	0.0598			0.28
Firm 176	200	0.0053	0.9947	0.0056	0.9944	0.0004			0.28
Firm 177	200	0.0053	1.0000	0.0056	1.0000	0.0000			0.28
Total	37780				Maximum D=	0.0626	0.0842	0.1053	95.96
					KSG	1.389			

Table 7. Simulation 37, with 177 firms.

Legend: Firm= “Name of Firm”; Production: quantity of units produced by the firm, for instance; w_i = observed share of each firm in the market; Aw_i = accumulated observed share; T_i = theoretical share of each firm in the case of equal production; At_i = accumulated theoretical production; D= difference between accumulated values: $Aw_i - T_i$; C4= concentration index CK with $k=4$ calculated in accordance with Formula 6; C8= likewise with $k=8$; HHI= Herfindahl-Hirschman index calculated in accordance with Formula 1. Maximum D: maximum observed D value in n cases; KSG= proposed KSG index calculated in accordance with Formula 11: **Source:** Author.

In the following sub-section, the hypotheses formulated above are tested.

6.2. Test of hypotheses

6.2.1 The results of testing H1

The analysis shows that for a given size n of market, the correlation between the KSG and HHI is perfect. This can be seen in the multiple graphs presented in **Figure 3**, presenting 100% correlations with R^2 . In other words, with the n number of firms in the market, there is a perfect correlation between Formulas 11 and 1:

$$KSG = \chi^2 = 4D^2 \frac{n_1 n_2}{n_1 + n_2} = 4D^2 \frac{n^2}{2n} = 2D^2 n \quad \text{and} \quad IHH = \sum_{i=1}^n 10000 \left(\frac{w_i}{w} \right)^2$$

There was a perfect correlation between the KSG and HHI values, to such an extent that it is possible to make a linear conversion from one value to the other, considering a given market with n firms. Therefore, hypothesis H₁ is not rejected: it could be said that there is no significant difference between the results of the KSG and HHI for a given n size market.

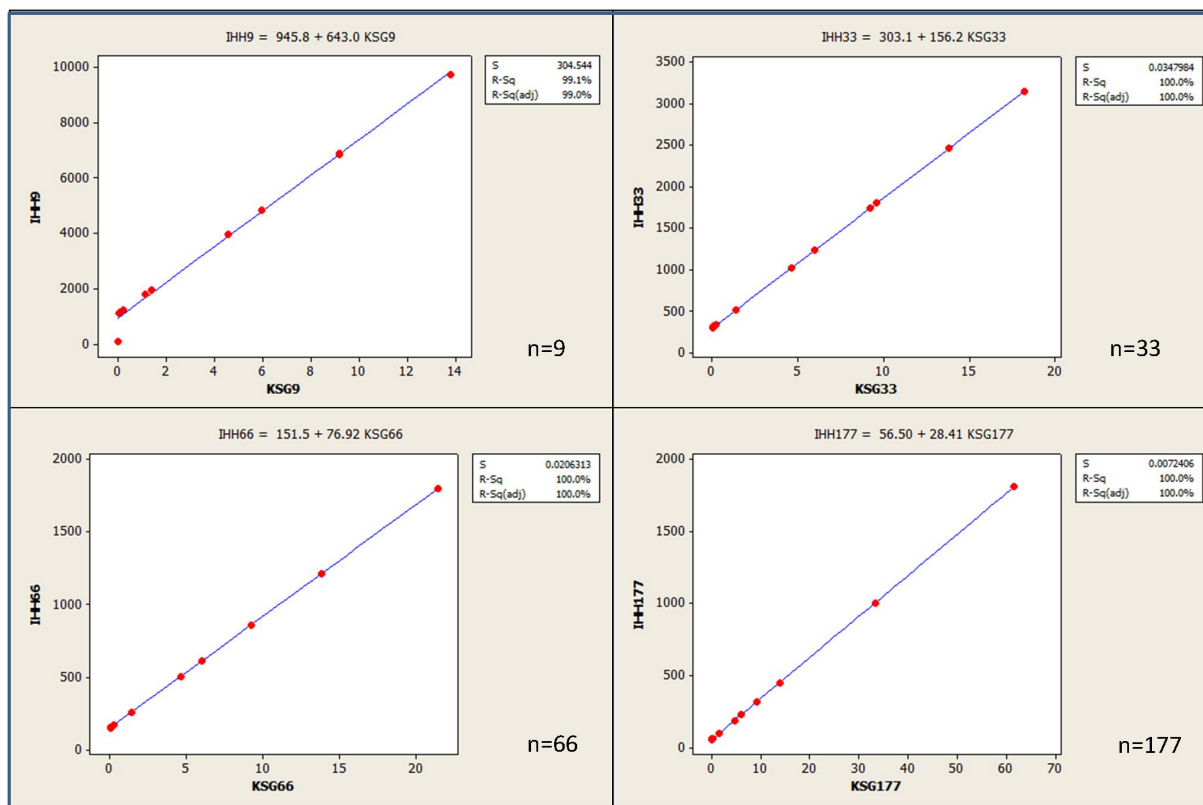


Figure 3; Multiple graph showing the associations between the HHI and KSG for several sizes of market (with 9, 33, 66 and 77 firms). *Source:* Author.

6.2.2. The results of testing H2

According to Table 1, the HHI value has a decisive interpretation two occasions: when HHI=1000, which expresses the passage from a concentrated market to a moderately concentrated market, and when HHI=1800, which shows the passage from a moderately concentrated market to a highly concentrated market (see Table 1).

The KSG has three general levels (free competition, imperfect competition and significant concentration) and each of these levels has three specific regions. The levels were established in accordance with the level of significance of χ^2 distribution. The market is considered significantly concentrated at a significance level of 0.05. The level of significance is the limit that is taken as a basis for affirming that a certain observed result is or is not a random result. Levels 0.05 and 0.01 are accepted as statistically significant, i.e., 5% and 1%, respectively. Above 0.05, the results observed are due to the law of randomness and the result is considered non-significant. The interpretation of the value of p is shown in Table 8 (Motulsky, p.9)^[40]. Strictly speaking, in the study of market concentration, it is only interesting when the level of concentration has a p value equal to or less than 0.05 – a concentrated market indicator.

Valor of p	Interpretation
>0.05	Non-significant
0.01 - 0.05	Significant
0.001 - 0.01	Very significant
<0.001	Extremely significant

Table 8. Interpretation of observed p value

Source: Motulsky (p.9)^[41]

The KSG can be associated with the χ^2 distribution and its value has an immediate statistical interpretation (Table 2), which does not occur with the HHI index. Although not relevant, it is possible to establish for the KSG a specific significance that enables the establishment of level of change of significance. At each of these levels, for the KSG and

interpretation of market concentration has been defined (Table 3). This type of interpretation is not possible for the HHI index. Thus, hypothesis H₂ is not rejected. The results shown by the KSG indicator have greater informational content than the results presented by the HHI.

7. Conclusions

According to Kon^[15], one of the most commonly used indexes for measuring the concentration of an industry is the HHI and in this study a new index is proposed that is based on the KS (Kolmogorov-Smirnov) test and has its interpretation of significance given by Goodman and Kruskal^[35]. For this reason, the index is designated as the KSG. The proposed model uses non-parametric techniques to establish the dimensioning of concentration and define the level of significance of the value found, and its application is equal to the application of the KS test. With the value obtained, consulting a table of χ^2 distribution with two degrees of freedom, it is determined whether the values observed differ significantly from the theoretical values that express a perfect market. The variable to investigate can be any one, but is typically the volume of production or sales or some other equivalent variable.

Forty-three simulations were made, using four different n values of firms in the market (9, 33, 66 and 177 firms) to test two hypotheses.

Starting from the principle that it is desirable to have concentration indicators to measure the level of concentration of a market and that these measures should not conflict with one another and taking into account that the HHI is one of the most used indexes, a hypothesis was set that the proposed KSG index has no significant difference in results in the analysis of a market with n firms.

The test of this hypothesis showed that

$$KSG = \chi^2 = 4D^2 \frac{n_1 n_2}{n_1 + n_2} = 4D^2 \frac{n^2}{2n} = 2D^2 n \quad \text{and} \quad IHH = \sum_{i=1}^n 10000 \left(\frac{w_i}{w} \right)^2$$

are linearly correspondent (see **Figure 3**). It can be said that for a given size of n market, the function observed is

$$IHH = a + b(KSG) \tag{13}$$

The observed a and b values in the regressions of the four markets (with 9, 33, 66 and 177 firms) and shown in **Figure 3** are in **Table 8**. The R² determination coefficient is practically 100%. The determination coefficient can also be defined as the degree of adjustment of the estimated straight line in the set of data, i.e., it shows how well the model adjusts to the set of data. In the present case, the adjustment is perfect, i.e., there is perfect adherence between the HHI and KSG.

n	a	B	R ²
9	945.8	643	99.1
33	303.1	156.2	100
66	151.5	76.92	100
177	56.5	28.41	100

Table 8. Observed a and b values

Source: Author

The second hypothesis sought to show that the quality of information of the KSG concentration index is greater than that given by the HHI. When the HHI index is calculated, what is basically obtained is information given by its interpretation (Table 1); when the KSG index is obtained, the level of significance is also the same. This information is particularly relevant when analyzing markets in cases of company mergers.

It is believed that the indicator proposed in this study is fully justified.

References

1. Lerner PA. The concept of monopoly and the measurement of monopoly power. *Review of Economic Studies* 1934; 1(3): 157-175.
2. Iwata G. Measurement of conjectural variation in oligopoly. *Econometrica* 1974; 42(5): 947-966.
3. Braga H, Mascolo J. Mensuração da concentração industrial no Brasil. *Pesquisa e Planejamento Econômico* 1982;

- 12(2): 399-454.
4. Tirole J. A teoria da organização industrial. Cambridge, Mas.: MIT Press 1988.
 5. Scherer FM, Ross D. Industrial market structure and economic performance. (3a ed.). Boston: Houghton Mifflin 1990.
 6. Tonge SD, Wooton CW. (1991). Auditor concentration and competition among the large public accounting firms: post-merger status and future implications. *Journal of Accounting and Public Policy* 1991; 10: 157-172.
 8. Baker JB, Bresnahan TF. Empirical methods of identifying and measuring market power. *Antitrust Law Journal* 1992; 61(1): 3-16.
 10. Cuevas F. La Reglamentacion de un Monopólio Natural: El Caso de La Industria Eléctrica en América Latina; Un Enfoque Político-Econômico. Tese de doutorado, Universidade de Montpellier I, Montpellier, França 1993.
 11. Tiffin AL, Dawson PJ. (1997). Measuring oligopolistic distortion in the UK frozen potato product sector: a calibration modeling approach. *Journal of Agricultural Economics* 1997; 48(3): 300-312.
 12. Borenstein S, Bushnell J, Knittel C. (1999). Market Power in Electricity Markets: Beyond Concentration Measures. *The Energy Journal* 1999; 20(4): 65-88.
 13. Mahajan S. Concentration ratios for businesses by industry in 2004. *Economic Trends* 2006; (635): 25-47.
 14. Ballas AA, Fafaliou I. Market shares and concentration in the EU auditing industry: The effects of andersen's demise. *International Advances in Economic Research* 2008; 14(4): 485-497.
 15. Kozyrev O, Malyzhenkov P. Industrial clusters as the form of the territorial organization of economy in Russia and Italy. *European Journal of Economics, Finance And Administrative Sciences* - 42 (2011); 133-138.
 16. Al-Jarrah IM, Qasrawi W, Obeidat BY, *et al.* Evaluating the competition and pricing power in the banking sector of Jordan. *European Journal of Economics, Finance and Administrative Sciences* - 46 (2012); 41-53.
 17. Kon A. *Economia Industrial*. São Paulo: Nobel 1994.
 18. Kupfer D, Hasenclever L. (Orgs). *Economia industrial: fundamentos teóricos e práticas no Brasil*. Rio de Janeiro: Campus 2002.
 19. Clarke R. *Industrial Economics*. Cambridge Massachusetts: Blackwell 1993.
 20. Tabner IT. The relationship between concentration and realised volatility: an empirical investigation of the FTSE 100. Doctoral thesis, University of Stirling, Stirling, 2003.
 21. Hirschman AO. *National Power and the Structure of Foreign Trade*. Berkeley: University of California Press 1945.
 22. Herfindahl OC. *Concentration in the US Steel Industry*. Doctoral Dissertation Unpublished, Columbia University, New York 1950.
 23. Simpson EH. Measurement of Diversity. *Nature* 1949; 163(13): 688.
 24. Yule GU. *Statistical Study of Literary Vocabulary*. Cambridge: Cambridge University Press 1944.
 25. Fisher RA, Corbet SA, Willams CB. The relation between the number of species and the number of individuals in a random sample of an animal population. *Journal of Animal Ecology* 1943; 12(1): 42-58.
 26. Williams CB. (1946). Yule's Characteristic and the Index of Diversity. *Nature* 1946; 157(13): 482.
 27. Hannah, Kay JA. *Concentration in Modern Industry. Theory, Measurement and the U.K. Experience*, London: The Macmillan Press 1977.
 28. Shannon CE. A mathematical theory of communication. *Bell System Technical Journal* 1948; 27(3): 379-423.
 29. Shannon CE, Weaver W. *The mathematical theory of communication* urbana and Chicago. University of Illinois Press 1946.
 30. Nyquist H. Certain factors affecting telegraph speed. *Bell System Technical Journal* 1924; 3(4): 324-346.
 31. Hartley RVL. Transmission of information. *Bell System Technical Journal* 1928; 7(4): 535-563.
 32. Wiener N. *Cybernetics, or control and communication in the animal and the machine*. New York and London: MIT Press 1961.
 33. Gini C. *Memorie di Metodologica Statistica, Variabilità e concentrazione*. Milano: Dott. A. Giuffrè 1939; 359-408.
 34. Lorenz MO. Methods of measuring the concentration of wealth. *Quarterly Publications of the American Statistical Association* 1905; 9 (70): 209 -219
 35. HMG- Horizontal Merger Guidelines, U.S. Department of Justice and the Federal - Trade Commission, 1997.
 36. Fernholz R, Garvy R, Hannon J. Diversity-Weighted Indexing. A New Approach to Passive Investing. *The Journal of Portfolio Management* 1998; 24(2): 74-82.
 37. Goodman LA, Kruskal WH. Measures of association for crossclassification. *J. Amer. Statist. Assoc.* 1954; 49: 732-764.
 38. Rubinfeld R. Robust functional equations and their applications to program testing. FOCS 1994.
 39. Baquero G. *Métodos de Pesquisa Pedagógica*. São Paulo: Loyola 1970.
 40. Siegel S. *Nonparametric Statistics for the Behavioral Sciences*. Tokyo: Kogakusha 1960.
 41. Smirnov N. Table for estimating the goodness of fit of empirical distributions. *Annals of Mathematical Statistics* 1948; 19: 279-281. doi:10.1214/aoms/1177730256

42. Motulsky HJ. Analyzing data with graphPad prism. San Diego (CA): GraphPad Software 1999.
43. Morettin PA, Bussab WO. Métodos Quantitativos (4a ed.). São Paulo: Atual 1991.