Bounds of Herfindahl-Hirschman index of banks in the European Union

Tóth, József

King Sigismund Business School, Hungary

4 February 2016

JÓZSEF TÓTH

Bounds of Herfindahl-Hirschman index of banks in the European Union

Abstract

The level of concentration can be measured by different indicators. The article overviews the calculation methods of the so called CR3 index, Gini-index and Herfindahl-Hirschman index. The paper makes evidence for existence of such upper and lower bounds of Herfindahl-Hirschman index that could be determined by sampling if the total market value and the total number of market players are known. Furthermore, it proves that the difference of the upper and lower bounds decreases if the sample is supplemented by new units of the analyzed population. In order to apply the findings, the bounds of the Herfindahl-Hirschman index of the bank market of the European Union are determined in case of total assets, own funds, net interest income and net fee income of European banks. It can be done because these aggregated values and the number of European banks are known.

Journal of Economic Literature (JEL): C01, C51, G21

Key Words: Herfindahl-Hirschman index, concentration, bank

1. Introduction

Measuring of the concentration level of the bank market is important for more reasons.

On the one hand, it is important because evaluation of the processes effecting on the risk level of the banking activity supports mitigation of the risks hidden in the market concentration. In 90's, the European banking market went through in a significant concentration process. As a result of it, huge banks have been created. For example, the balance sheet total of the three largest banks in the United Kingdom is close to two and half times higher than the GDP of the UK in 2014. The situation of other European Union member states is the same. It bears extreme high risk. Bankruptcy of a large bank might cause unmanageable consequences. Therefore, measuring the concentration level and making decision on this result are important in the modern banking industry. Generally, the financial governments have tools to measure concentration via national reporting systems.

On the other hand, determination the strength of the competition supports the decisions of the market players. If a bank wants to enter a new market, measuring of the competition level is an unavoidable task. Also, a bank operating within a market should analyse the possibilities of gaining new market share. Since, the competition level depends on the concentration level and vice versa, the new and old market players should measure the concentration of the market.

Even more, when making decisions, investors of banking industry also have to have information on banking market concentration level.

Since survey of the entire market is extremely expensive and cannot be done in numerous cases, sampling the market is the acceptable way to get information about the market concentration. Though different methods have already been elaborated (Nauenberg, Basu and Chan, Naldi and Flamini, Michelinia and Pickforda) , their application is not always easily feasible in case of financial industry. However, since the banking is a special market, numerous data are reported to the European and national central banks. If these data could be used, it would make the calculation of the concentration level easier.

The main goal of this paper is to introduce such tool that applicable for market players and decision makers for determination of the banking market concentration in a simple way based on data on banking activities disclosed by the European Central Bank and based on sampling.

The applied indicator is the Herfindahl-Hirschman index (HHI).

2. Calculation of Herfindahl-Hirschman index

The concentration level can be measured based on balance sheet total, net interest income, number of clients etc. This paper uses balance sheet and profit or loss statement data.

The formula of HHI is calculated as follows: $HHI = \sum_{i=1}^{n} \left(\frac{x_i}{T} \cdot 100\right)^2$ or $HHI = \sum_{i=1}^{n} \left(\frac{x_i}{T}\right)^2$, where

n is the number of the market players,

x_i is the value of the *i*th market player,

$\mathrm{T} = \sum_{\mathrm{i}=1}^{\mathrm{n}} \mathrm{x_i}$, and $0 < \mathrm{i} \leq \mathrm{n}, \mathrm{i} \in \mathbb{Z}^+, \mathrm{n} \in \mathbb{Z}^+$ (when calculating the concrete values bellow, figures are given in percentage form but in other cases in normal form).

This index is easily can be calculated if all of the measured value of the market players is known. For example, if the market consists of four market players, the analysed market concentration value is the total assets of the market players and their values are {120;200;80;500}. In that case the sum of the total assets is 900 and the Herfindahl-Hirschman index is the following:

$$HHI = \left(\frac{120}{900} \cdot 100\right)^2 + \left(\frac{200}{900} \cdot 100\right)^2 + \left(\frac{80}{900} \cdot 100\right)^2 + \left(\frac{500}{900} \cdot 100\right)^2 = 3837.037$$

The HHI values are between zero and 10000. The higher level of the concentration is expressed by higher level of the index.

If the market share of the market players is not known and numerous market players operate in the market, determination of the HHI value is not so easy.

If there is no possibility to get information about all of the market shares the Herfindahl-Hirschman index is to be estimated.

This paper aims to make prove for the following two allegations:

a. If the number of market players of the whole banking market and aggregated value of their balance sheet or profit or loss statement items are known, bounds of Herfindahl-Hirschman index are determinable based on sampling in connection with the concentration of the balance sheet or profit or loss statement items in question.

b. By increasing the sample size, the interval determined by bounds specified above decreases. In other words the estimation of the Herfindahl-Hirschman index is more accurate if the sample size increases.

Though the total assets related HHI index is also disclosed by the European Central Bank, but when calculating, insurance institutions and pension funds are also included. Using the elaborated method introduced bellow, any balance sheet data or any profit or loss statement data concentration in relation with financial institutions can be estimated.

3. Former Herfindahl-Hirschman index estimation methods

When estimating the index, the base condition of Nauenberg, Basu and Chan (1997) was that the market share of the largest market players is known but other's market share is unknown. They analysed the likely distribution of the unknown market participants' market share. In their approach the market share of the unknown market participants is given in a positive integer value which is expressed in percentage. They originated the issue from the well-known combinatorial problem where m pieces ball should be distributed in q boxes and all of the balls must be ordered to one of the boxes as well as each box should be filled up at least one ball. In this problem the m means the market share given as an integer within a market which must be distributed among *q* market players.

According to the example of the authors, if the market has *n* market players and the market share of the three smallest firms (q=3) is 8 % (q=8%) than the following distribution and probability can be expressed by permutation:

Table 1: Example for Neuenberg et al.'s forecast of the Herfindahl-Hirschman index

Number	Permutations			Sum	HHI value	Probability	Expected value
	X_{n-2}	X_{n-1}	X_n	m			
1	1	1	6	8			
2	1	6	1	8			
3	6	1	1	8	38	3/21=0,14	5,43
4	1	2	5	8			
5	1	5	2	8			
6	2	1	5	8			
7	2	5	1	8			
8	5	1	2	8			
9	5	2	1	8	30	6/21=0,29	8,57
10	1	3	4	8			
11	1	4	3	8			
12	3	1	4	8			
13	3	4	1	8			
14	4	1	3	8			
15	4	3	1	8	26	6/21=0,29	7,43
16	2	2	4	8			
17	2	4	2	8			
18	4	2	2	8	24	3/21=0,14	3,43
19	2	3	3	8			
20	3	2	3	8			
21	3	3	2	8	22	3/21=0,14	3,14
Sum							28

Source: Based on data of Neuenberg et al., own editing

In the example above, the expected HHI value of the unknown market shares is 28. Therefore, the HHI of the whole market is the HHI of the known shares of the market players enlarged by 28.

Also, Naldi and Flamini (2014) made interval forecast for the Herfindahl-Hirschman index. In their approach, the market share of the *M* largest market players is also known but other's market share is unknown. $HHI = \sum_{i=1}^{n} s_i^2$ as well as $1 = \sum_{i=1}^{n} s_i$, if the s_i expresses the market share of the *i*th market player in the market where there are *n* market participants and the market share of the *M* largest market players is known. In that case Naldi and Flamini determined the minimal value of the HHI as follows: $HHI_{min} = \sum_{i=1}^{M} s_i^2 + \frac{\left(1-\sum_{i=1}^{M} s_i\right)^2}{n-M}$. When determining the maximal value of the Herfindahl-Hirschman index, they differentiated two cases. If the aggregated value of the unknown market share is less than the least known market share than $HHI_{max} = \sum_{i=1}^{M} s_i^2 + \left(1-\sum_{i=1}^{M} s_i\right)^2$.

If the aggregated value of the unknown market share is higher than the least known market share than

$HHI_{max} = \sum_{i=1}^{M} s_i^2 + s_M^2 + (1-\sum_{i=1}^{M} s_i - s_M \cdot Q)^2$, where $Q = \frac{1-\sum_{i=1}^{M} s_i}{s_M}$.

Furthermore, Michelinia and Pickforda (1985) applied the concentration ratios in order to determine the lower and upper bounds of Herfindahl-Hirschman index, where the income of the enterprises was the analysed market value.

4. The new model

During the model compilation it is supposed that the aggregated value analysed and the number of units of the population is known. Furthermore, it is supposed that result of a random sample from the given population is also available.

Proposition 1

Let $0 < x_i$ denote the value of the units of a population which has $n \in \mathbb{Z}^+$ units where $i \leq n$ and $i \in \mathbb{Z}^+$. Let the sample size originated from the population is $\mathrm{n} - \mathrm{k}$ $(1 < \mathrm{k} \leq \mathrm{n},\ \mathrm{k} \in \mathbb{Z}^+)$ and $\bar{x} = \frac{G}{k}$, where $G = \sum_{i=1}^{k} x_i$ denotes the mean of the unknown units as well as let at least one unknown unit value which in not equal to at least one other unknown unit value. In that case, by substituting each unknown unit by their mean, the HHI value of the population composed by $n - \mathrm{k}$ known units and k pieces $\bar{x}$ is lower than the HHI of the original population.

If the proposition is true, the lower bound level of the HHI is determinable by sampling in case where the aggregated value and the number of units of the population are known. It means that knowing the aggregated value of a population (for example the aggregated value of the total assets of banks funded on the area of European Union) and the number of units of a population (the number of those banks), than a lower bound determinable (by calculation of the mean of the total assets of the unknown banks and substituting each unknown total asset value by their mean). In this case the HHI value of the population is higher than the lower bound.

For example let the A the set of the following values: A= {5;10;20;25;40;50;60;80;100}. A has 9 units and sum of the elements is 390. The HHI value of the population is

$$HHI = \left(\frac{5}{390} \cdot 100\right)^2 + \left(\frac{10}{390} \cdot 100\right)^2 + \left(\frac{20}{390} \cdot 100\right)^2 + \left(\frac{25}{390} \cdot 100\right)^2 +$$
$$+ \left(\frac{40}{390} \cdot 100\right)^2 + \left(\frac{50}{390} \cdot 100\right)^2 + \left(\frac{60}{390} \cdot 100\right)^2 + \left(\frac{80}{390} \cdot 100\right)^2 + \left(\frac{100}{390} \cdot 100\right)^2 =$$
$$= 1660.09.$$

If the sample has 4 units than 5 units is unknown. For example let the sample is the following: {25;50;80;100}. Therefore {5;10;20;40;60} are unknown. The essence of the method is that the unknown units are substituted by their mean during the HHI calculation. Though, {5;10;20;40;60} are unknown, their mean is determinable because according to the original assumption, sum of the value of the units and number of the units is known. Since sum of the units is 390, sum of the known elements is 25+50+80+100=255 and the number of the unknown units is 5, therefore, the mean of the unknown units is $\frac{390-255}{5} = 27$. Let *A'* the population of the known values and 5 pieces 27: *A'*= {27;27;27;25;27;50;27;80;100}. In that case the HHI value of the modified population is the following:

$$\overline{HHI} = \left(\frac{27}{390} \cdot 100\right)^2 + \left(\frac{27}{390} \cdot 100\right)^2 + \left(\frac{27}{390} \cdot 100\right)^2 + \left(\frac{25}{390} \cdot 100\right)^2 + \\ + \left(\frac{27}{390} \cdot 100\right)^2 + \left(\frac{50}{390} \cdot 100\right)^2 + \left(\frac{27}{390} \cdot 100\right)^2 + \left(\frac{80}{390} \cdot 100\right)^2 + \\ + \left(\frac{100}{390} \cdot 100\right)^2 = 1523.34 .$$

By substituting the 5 unknown elements by their mean, such value was determinable which is lower than the HHI of the original population. In this case the HHI value is higher than 1523.34.

Beside the denotations in the Proposition 1, let sum of the value of the units of the population $T = \sum_{i=1}^{n} x_i$. Let denote $\overline{HHI}$ the Herfindahl-Hirschman index of such population that composed by $n -$ k known units and k pieces $\overline{x}$. The order of the units does not have effect on HHI value, therefore, let the unknown k elements are the first units $(x_1, x_2, x_{3,\dots,} x_k)$, and $x_{k+1}, x_{k+2}, x_{k+3,\dots,} x_n$ the known elements of the modified population. In that case, $\overline{HHI} < HHI$ statement must be proven.

Since $\frac{1}{k}=\frac{x_1}{G}=\frac{x_2}{G}=\frac{x_3}{G}=\cdots=\frac{x_k}{G}$ is not possible (there is at least one unknown unit value which in not equal to at least one other unknown unit value), therefore, the value of the expression $\left(\frac{1}{k}-\frac{x_1}{G}\right)^2+\left(\frac{1}{k}-\frac{x_2}{G}\right)^2+\left(\frac{1}{k}-\frac{x_3}{G}\right)^2+\cdots+\left(\frac{1}{k}-\frac{x_k}{G}\right)^2$ is higher than zero. After leaving the brackets $0<\sum_{i=1}^{k}\frac{1}{k^2}-\frac{2\cdot x_i}{k\cdot G}+\left(\frac{x_i}{G}\right)^2$. Therefore, $0<\frac{1}{k}-\frac{2}{k}\sum_{i=1}^{k}\frac{x_i}{G}+\sum_{i=1}^{k}\left(\frac{x_i}{G}\right)^2$. Since $\sum_{i=1}^{k}\frac{x_i}{G}=1$, thus,

$0<\frac{1}{k}-\frac{2}{k}\cdot 1+\sum_{i=1}^{k}\left(\frac{x_i}{G}\right)^2$. As $\frac{1}{k}-\frac{2}{k}=-\frac{1}{k}$. Accordingly, $\frac{1}{k}<\frac{\sum_{i=1}^{k}{x_i}^2}{G^2}$. It means $k\cdot\left(\frac{G}{k}\right)^2<\sum_{i=1}^{k}{x_i}^2$. Multiplying both side of the inequality by $\left(\frac{1}{T}\right)^2$,

$k\cdot\left(\frac{\frac{G}{k}}{T}\right)^2<\sum_{i=1}^{k}\left(\frac{x_i}{T}\right)^2$, where $\frac{G}{k}$ the mean of the first k elements. Increasing both side of the inequality by $\sum_{i=k+1}^{n}\left(\frac{x_i}{T}\right)^2$ the inequality will be the following: $k\cdot\left(\frac{\frac{G}{k}}{T}\right)^2+\sum_{i=k+1}^{n}\left(\frac{x_i}{T}\right)^2<\sum_{i=1}^{k}\left(\frac{x_i}{T}\right)^2+\sum_{i=k+1}^{n}\left(\frac{x_i}{T}\right)^2$. Therefore,

$k\cdot\left(\frac{\frac{G}{k}}{T}\right)^2+\sum_{i=k+1}^{n}\left(\frac{x_i}{T}\right)^2<\sum_{i=1}^{n}\left(\frac{x_i}{T}\right)^2$. Thus, $\overline{HHI}\leq HHI.$ It is exactly the statement given in the proposition.

When calculating the concentration of the bank market of the European Union, the mean of the values of unknown units is different in case of different member states. Since in case of any k units (where 1< $k\leq n$), substituting the values of units by their mean the value of HHI decreases, therefore, the Proposition 1 makes possible that the mean of the unknown values are to be calculated as per member states. By using mean of unknown values as per member states, the value of HHI decreases comparing with HHI value of the original population.

<u>Proposition 2</u>

Let $0 < x_i$ denote the value of the units of a population which has $n \in \mathbb{Z}^+$ units where $i \leq n$ and $i \in \mathbb{Z}^+$. Let the sample size originated from the population is $\mathrm{n} - \mathrm{k}$ $(1 < \mathrm{k} \leq \mathrm{n},\ \mathrm{k} \in \mathbb{Z}^+)$ where $G = \sum_{i=1}^{k} x_i$ denotes the sum of the unknown units. By excluding the unknown units from the population and supplementing it by G, the Herfindahl-Hirschman index value of the population created on this way (it has $n - k + 1$ units) is higher than the Herfindahl-Hirschman index value of the original population (which has n units).

Keeping the above example, A= {5;10;20;25;40;50;60;80;100}. It has nine units, the sum of the value of the units is 390, the value of the HHI is 1660.09. The sample is {25;50;80;100}. The essence of the method is that the unknown units are excluded and the population is supplemented by sum of their values. The HHI of the population created on this way is higher than the HHI value of the original population. Since sum of the units of A is 390 and sum of the value of the units in the sample 25+50+80+100=255, therefore, G= $390 - 255 = 135$. According to the proposition, the HHI of the A population is less than the HHI of the population composed by {25;50;80;100;135} units ($\overline{HHI}$). It is less actually, since

$$\overline{HHI} = \left(\frac{25}{390} \cdot 100\right)^2 + \left(\frac{50}{390} \cdot 100\right)^2 + \left(\frac{80}{390} \cdot 100\right)^2 +$$

$$+ \left(\frac{100}{390} \cdot 100\right)^2 + \left(\frac{135}{390} \cdot 100\right)^2 = 2481.92.$$

By excluding the 5 unknown elements and supplementing the population by their aggregated value, such value was determinable which is higher than the HHI of the original population. In this case the HHI value is less than 2481.92.

Beside the denotations in the Proposition 2, let sum of the value of the units of the population $T = \sum_{i=1}^{n} x_i$. Let denote $\overline{HHI}$ the Herfindahl-Hirschman index of such population that composed by $n - \mathrm{k} + 1$ ($\{G, x_{k+1}, x_{k+2}, x_{k+3,\dots,} x_n\}$) units. In that case, $HHI < \overline{HHI}$ statement must be proven.

Since $0 < x_i$ in any case,

$$0 < 2 \cdot (x_1 \cdot x_2 + x_1 \cdot x_3 + \cdots + x_1 \cdot x_n + x_2 \cdot x_3 + x_2 \cdot x_4 + \cdots + x_2 \cdot x_n + \cdots + x_{k-1} \cdot x_k).$$

Increasing the right side of inequality by $0 = \sum_{i=1}^{k} x_i^2 - \sum_{i=1}^{k} x_i^2$,

$$0 < (x_1 + x_2 + x_3 + \cdots + x_k)(x_1 + x_2 + x_3 + \cdots + x_k) - \sum_{i=1}^{k} x_i^2.$$

Therefore, $0 < G^2 - \sum_{i=1}^{k} x_i^2$ and $\sum_{i=1}^{k} x_i^2 < G^2$. Multiplying both side of the inequality by $\left(\frac{1}{T}\right)^2$, $\sum_{i=1}^{k} \left(\frac{x_i}{T}\right)^2 < \left(\frac{G}{T}\right)^2$. Increasing both side of the inequality by $\sum_{i=k+1}^{n} \left(\frac{x_i}{T}\right)^2$ the inequality will be the following:

$$\sum_{i=1}^{k} \left(\frac{x_i}{T}\right)^2 + \sum_{i=k+1}^{n} \left(\frac{x_i}{T}\right)^2 < \left(\frac{G}{T}\right)^2 + \sum_{i=k+1}^{n} \left(\frac{x_i}{T}\right)^2.$$

Therefore, $HHI < \overline{HHI}$. It is exactly the statement given in the proposition.

When calculating the concentration of the bank market of the European Union,, the sum of the values of unknown units is different in case of different member states. Since in case of any k units (where 1< $k \leq n$) the HHI value of the original population is less than the $\overline{HHI}$, the Proposition 2 makes possible that the sum of the unknown values are to be calculated as per member states. By using sum of the unknown values as per member states, the value of HHI increases comparing with HHI value of the original population.

Proposition 3

Supplementing the sample size by an additional data collection, the value of HHI of the modified population determined in Proposition 1 will increase.

In the example used at Proposition 1 the sample size was 4. Let the sample size now 5 where the known values are {25;40;50;80;100}. Sum of their values is 25+40+50+80+100= 295. The mean of the unknown values is $\frac{390-295}{4} = 23.75$. Therefore, the value of $\overline{HHI}$:

$$\overline{HHI} = \left(\frac{23{,}75}{390}\cdot 100\right)^2 + \left(\frac{23{,}75}{390}\cdot 100\right)^2 + \left(\frac{23{,}75}{390}\cdot 100\right)^2 + \left(\frac{25}{390}\cdot 100\right)^2 +$$

$$+ \left(\frac{40}{390}\cdot 100\right)^2 + \left(\frac{50}{390}\cdot 100\right)^2 + \left(\frac{23{,}75}{390}\cdot 100\right)^2 + \left(\frac{80}{390}\cdot 100\right)^2 + \left(\frac{100}{390}\cdot 100\right)^2 = 1537.23$$

When four units were known 1523.34 was the value of the HHI (which was less than the HHI of the original population). By adding new data to the sample the HHI value (which was also less than the original HHI) increased (1537.23). Therefore, the accuracy of the forecast is better.

Let $HHI_x = \sum_{i=1}^{n}\left(\frac{x_i}{T}\right)^2$ and $HHI_y = \sum_{i=1}^{n}\left(\frac{y_i}{T}\right)^2$ the Herfindahl-Hirschman indexes of the populations where *n* is the number of units in both cases. Let $T = \sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i$ where $i \le n, i \in \mathbb{Z}^+, n \in \mathbb{Z}^+$ and let $HHI_x < HHI_y$. Let supplemented both population by units $z_{n+1}, z_{n+2}, z_{n+3}, \dots z_{n+l}$. In that case $HHI'_x = \sum_{i=1}^{n}\left(\frac{x_i}{T'}\right)^2 + \sum_{i=n+1}^{n+l}\left(\frac{z_i}{T'}\right)^2$ and

$HHI'_y = \sum_{i=1}^{n}\left(\frac{y_i}{T'}\right)^2 + \sum_{i=n+1}^{n+l}\left(\frac{z_i}{T'}\right)^2$ where T' is sum of the values of the populations. Multiplying both side of the inequality $\sum_{i=1}^{n}\left(\frac{x_i}{T}\right)^2 < \sum_{i=1}^{n}\left(\frac{y_i}{T}\right)^2$ by $\left(\frac{T}{T'}\right)^2$ the new expression is $\sum_{i=1}^{n}\left(\frac{x_i}{T'}\right)^2 < \sum_{i=1}^{n}\left(\frac{y_i}{T'}\right)^2$.

Increasing both side of the inequality by $\sum_{i=n+1}^{n+l}\left(\frac{z_i}{T'}\right)^2$,

$\sum_{i=1}^{n}\left(\frac{x_i}{T'}\right)^2 + \sum_{i=n+1}^{n+l}\left(\frac{z_i}{T'}\right)^2 < \sum_{i=1}^{n}\left(\frac{y_i}{T'}\right)^2 + \sum_{i=n+1}^{n+l}\left(\frac{z_i}{T'}\right)^2$.

Therefore, $HHI'_x < HHI'_y$.

In other words if number of units and sum of the units of two populations is the same and the HHI value of the first population is less than the HHI value of the second one as well as both of the populations are supplemented with the same new units, relation of the HHI values does not change (the less remains the less).

Using the denotations above, let the first k units of the population the k unknown units $(x_1, x_2, x_{3,\ldots,}x_k)$, and $x_{k+1}, x_{k+2}, x_{k+3,\ldots,}x_n$ the known units where $\bar{x} = \frac{G}{k}$. By additional data collection let $x_1, x_2, x_3, \ldots, x_l$ $(l < k)$ are known. According to the Proposition 1, the Herfindahl-Hirschman index of the subpopulation (HHI_k) which has k units and composed by $x_1, x_2, x_3, \ldots, x_l$ known units and $k - l$ unknown units (their value is the mean of the unknown units) is higher than Herfindahl-Hirschman index of such subpopulation (HHI_{k-l}) that has k units (their value is $\bar{x}$): $HHI_k < HHI_{k-l}$. That is, the HHI value is lower if fewer units are known. Since the number of units is *k,* sum of the units is *G* in both of the subpopulations as well as the HHI value is lower in case of subpopulation where fewer units are known, supplementing the subpopulations by $x_{k+1}, x_{k+2}, x_{k+3,\ldots,}x_n$ known units, the HHI value will be lower where fewer units are known.

Proposition 4

Supplementing the sample size by an additional data collection, the value of HHI of the modified population determined in Proposition 2 will decrease.

In the example used at Proposition 2 the sample size was 4. Let the sample size now 5 where the known values are {25;40;50;80;100}. Sum of their values is 25+40+50+80+100= 295. The sum of the unknown values is $390 - 295 = 95$. Therefore the value of the $\overline{HHI}$:

$$\overline{HHI} = \left(\frac{25}{390} \cdot 100\right)^2 + \left(\frac{40}{390} \cdot 100\right) + \left(\frac{50}{390} \cdot 100\right)^2 + \left(\frac{80}{390} \cdot 100\right)^2 + \\ + \left(\frac{100}{390} \cdot 100\right)^2 + \left(\frac{95}{390} \cdot 100\right)^2 = 1982.25.$$

When four units were known 2481.92 was the value of the HHI (which was higher than the HHI of the original population). By adding new data to the sample the value (which is also higher than the original HHI) decreased (1982.25). Therefore, the accuracy of the forecast is better.

Using the denotations above, let the first k unknown units of the population denoted by $x_1, x_2, x_{3,\dots,} x_k$, and the known units denoted by $x_{k+1}, x_{k+2}, x_{k+3,\dots,} x_n$ where $G = \sum_{i=1}^{k} x_i$. By additional data collection let $x_1, x_2, x_3, \dots, x_l$ $(l < k)$ are known and let denote J the sum of such units that are not known after the additional (second) data collection ($J = \sum_{i=l+1}^{k} x_i$). Therefore, $x_1 + x_2 + \cdots + x_l + J = G$. According to the Proposition 4 the Herfindahl-Hirschman index of the population (HHI_{k-l}) composed by units $x_1, x_2, x_3, \dots, x_l, J, x_{k+1}, x_{k+2}, x_{k+3,\dots,} x_n$ is less than Herfindahl-Hirschman index (HHI_k) of the population composed by units $G, x_{k+1}, x_{k+2}, x_{k+3,\dots,} x_n$: $HHI_{k-l} < HHI_k$.

$x_1^2 + x_2^2 + \cdots + x_l^2 + J^2 < (x_1 + x_2 + \cdots + x_l + J)^2$, since $0 < x_i$ in any case. It means that $x_1^2 + x_2^2 + \cdots + x_l^2 + J^2 < G^2$. Therefore,

$$\left(\frac{x_1}{T}\right)^2 + \left(\frac{x_2}{T}\right)^2 + \cdots + \left(\frac{x_l}{T}\right)^2 + \left(\frac{J}{T}\right)^2 + \left(\frac{x_{k+1}}{T}\right)^2 + \left(\frac{x_{k+2}}{T}\right)^2 + \cdots + \left(\frac{x_n}{T}\right)^2 < \left(\frac{G}{T}\right)^2 + \left(\frac{x_{k+1}}{T}\right)^2 + \left(\frac{x_{k+2}}{T}\right)^2 + \cdots + \left(\frac{x_n}{T}\right)^2$$

In other words: $HHI_{k-l} < HHI_k$.

5. Practical application of method

In order to use the theoretical result above, a sample has been selected from the bank market of European Union. During the selection the following aspects were taken into account:

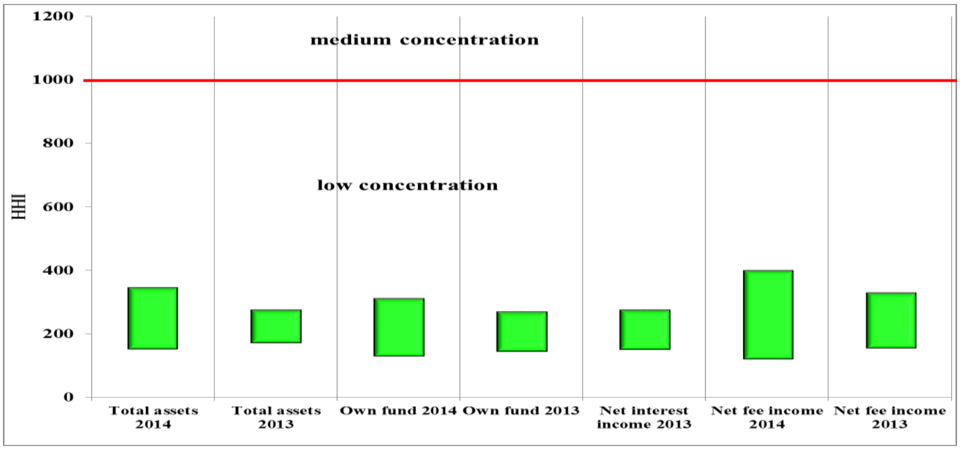
- at least five banks as per member states of European Union had to be covered by sampling,

- at least one to third of the total assets had to be covered by sampling as per member states.

Based on these aspects 164 banks were selected (the average value is 5,86 as per member states). The sample covers 72.40% of the whole bank market of the European Union (as per assets).

The lower and upper bounds were determined in case of total assets, own funds, net interest income and net fee income. The following chart depicts the lower and upper bounds of the analysed balance sheet or income statement data (data of net interest income are incomplete for 2014, these data are not indicated).

Chart 1: HHI bounds of total assets, own funds, net interest income and net fee income in the banking market of European Union

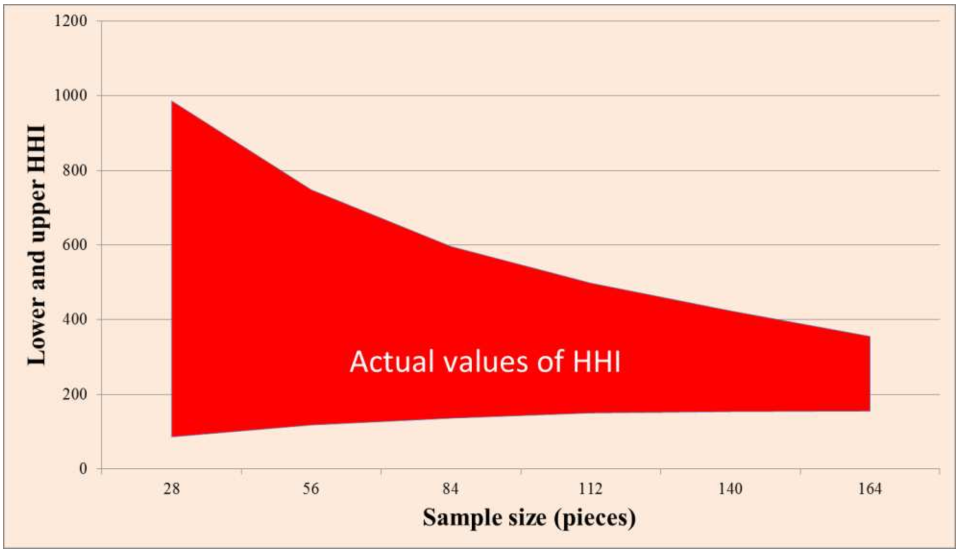


Source: Own calculation

The concentration level of the total assets, own funds, net interest income and net fee income was determinable by sampling. By using this method the cost of concentration measurement could be cut. However, it can be used only if the analysed aggregated value of the population as well as the number of market players is known. In case of European banks, the method is applicable because the European Central Bank discloses these data.

The lower and upper values of HHIs make possible to determine the concentration level by sampling. If the size of the sample increases the accuracy is better. It is observable if the lower and upper values are ordered to the sample size. For example, the following chart depicts how difference of upper and lower HHI values approaches each other when the sample size grows. At first, 28 banks are in the sample (from each member state). After that, the sample increases by 28 new banks (from each member state) and the difference between the upper and lower HHIs decreases.

Chart 2 : Change in the lower and upper HHI values by increasing the sample size.



Source: Own calculation

If the calculation relates to the separated market of member states, the concentration level increases. Thought, the sample size is not too high (the average is 5.86 per member state) the difference between the upper and lower HHI bounds is under 500 points in numerous cases. The following table summarizes the differences measured (data of net interest income are incomplete as for year 2014).

Table 2: Differences between upper and lower HHI bounds in case of 28 member states of European Union (number of member states, pieces)

Differences between the upper and lower HHIs (point)	**Total assets 2014**	**Total assets 2013**	**Own fund 2014**	**Own fund 2013**	**Net interest income 2014**	**Net interest income 2013**	**Net fee income 2014**	**Net fee income 2013**
0-100	5	4	3	4	1	1	3	4
100,1-200	1	1	3	1	3	3	1	1
200,1-300	1	1	2	2	4	4	3	2
300,1-400	0	2	3	2	3	1	1	2
400,1-500	2	1	0	3	2	2	2	1
500,1-600	1	1	2	1	2	3	2	2

Differences between the upper and lower HHIs (point)	Total assets 2014	Total assets 2013	Own fund 2014	Own fund 2013	Net interest income 2014	Net interest income 2013	Net fee income 2014	Net fee income 2013
600,1-700	2	2	2	2	1	3	3	5
700,1-800	0	1	1	1	0	1	1	1
800,1-900	1	3	0	2	3	2	1	1
900,1-1000	3	1	0	2	1	0	0	0
1000,1-	12	11	12	8	7	8	11	9
Sum	**28**	**28**	**28**	**28**	**27**	**28**	**28**	**28**

Source: own calculation

The most accurate measurement was in case of net interest income regarding year 2014 where close to half of the cases less than 500 points was the difference.

When analysing the applicability of the propositions it can be determined that how many cases was the concentration level unambiguously classified. The following table summarises the result.

Table 3: Classification based on the upper and lower HHI bounds

Cases	Total assets 2014	Total assets 2013	Own fund 2014	Own fund 2013	Net interest income 2014	Net interest income 2013	Net fee income 2014	Net fee income 2013
The category can be unambiguously determined	9	10	12	9	11	10	12	15
The upper HHI bound is in one category higher than the lower one	11	9	7	12	11	13	10	8

Cases	Total assets 2014	Total assets 2013	Own fund 2014	Own fund 2013	Net interest income 2014	Net interest income 2013	Net fee income 2014	Net fee income 2013
The upper HHI bound is higher than 1800 and the lower is less than 1000	8	9	9	7	5	5	6	5
Sum	28	28	28	28	27	28	28	28

Source: own calculation

In more than half of the cases the upper and lower HHI bound either in the same concentration category or the upper HHI bound is in one category higher than the lower one.

6. Conclusion

Different methods are elaborated for measuring the concentration level of the market. The Herfindahl-Hirschman index is one of the most applied market concentration measurement indexes. Since the actual market shares of the market players are not known in the practice, the Herfindahl-Hirchman index is to be estimated. Beside other previously elaborated methods, upper and lower bounds of Herfindahl-Hirschman index could be determined based on a sample originated from the entire population if the aggregated value of the analysed market is known. It is the case when analysing the European bank market concentration. Since, the European Central Bank publishes the aggregated balance sheet and profit or loss statement items and the number of the market participants, the method introduced above is applicable. Using these data, lower and upper bounds can be determined based on sampling.

The difference of the upper and lower bounds decreases if the sample is supplemented by new units of the analysed population. The practical application of the method shows also the same result.

The method introduce above makes possible for the market players to determine lower and upper bounds of the HHI index in a new way in connection with market of a member state of the European Union. Selecting 6-8 bigger market players' financial statements operating in the territory of the member state of the European Union and using their data as well as using data of the European Central Bank, the country specific concentration level can be estimated.

References:

1. European Central Bank (2015). Report on financial structures. European Central Bank

2. European Central Bank (2016). Statistical wirehouse

3. Fleming, J., Nellis, G. (2000). Principles of Applied Statistics. Thomson Learning, 2nd edition, p. 27

4. Herfindahl, O. C. (1950). Concentration in the U.S. steel industry. Unpublished doctoral dissertation, Columbia University

5. Michelinia, C.,Pickford, M. (1985). Estimating the Herfindahl Index from Concentration Ratio Data. Journal of the American Statistical Association, vol. 80 issue 390, pp. 301-305

6. Naldi, M., Flamini, M. (2014). Interval Estimation of the Herfindahl-Hirschman Index Under Incomplete Market Information. IEEE Computer Society, pp. 318-323

7. Nauenberg, E., Basu, K., Chand, H. (1997). Hirschman-Herfindahl index determination under incomplete information. Applied Economics Letters, vol. 4 issue 10, pp. 639-642