Chapter 3
Additional Indexes and Indicators for Assessment of Research Production

Dedicated to the Max-Planck Society (a treasure for scientific information ensuring high-quality research) and to the MPIPKS (one of the places where the quality of research work of young researchers has increased enormously in a short time)

Abstract About forty-five indexes for assessment of research production of single researchers have been discussed in Chap. 2. These indexes are based mainly on citations of publications of the evaluated researcher. The indexes form Chap. 2 can be calculated also for groups of researchers. In addition to indexes from Chap. 2, other indexes useful for assessment of production of groups of researchers may be used. About ninety such indexes are discussed in this chapter. The indexes are grouped in the following classes: simple indexes; indexes for deviation from simple tendency; indexes for difference; indexes for concentration, dissimilarity, coherence, and diversity; indexes for advantage and inequality; indexes for stratified data; indexes for imbalance and fragmentation; indexes based on the concept of entropy; Lorenz curve and associated indexes. In addition, the set of indexes connected to the RELEV method for assessment of scientific research performance within public institutes as well as indicators and indexes for scientific research performance of nations and about comparing national scientific productions are discussed. Finally, we discuss briefly several journal citation measures as well as an example of an application of a geometric tool for detection of scientific elites in a group of institutes on the basis of Lorenz curves.

3.1 Introductory Remarks

Two modes of knowledge production may be considered [1, 2]: Mode 1 and Mode 2. Mode 1 of knowledge production is motivated by scientific knowledge alone, e.g., by fundamental research. In other words, Mode 1 of knowledge production is not connected to the search for applications of the obtained results. Mode 1 of knowledge

N.K. Vitanov, *Science Dynamics and Research Production*, Qualitative and Quantitative Analysis of Scientific and Scholarly Communication,
DOI 10.1007/978-3-319-41631-1_3

production is founded on the separation of science into discrete disciplines (e.g., a researcher from one discipline may not bother about another discipline). In Mode 2 of knowledge production, multidisciplinary teams are brought together for short periods of time to work on specific problems in the real world for knowledge production. Mode 2 is closely connected to the project system of research, e.g., to how research funds are distributed among scientists and how scientists focus on obtaining these funds. In the case of Mode 1, the scientific knowledge production is carried out by actors who are distributed, yet proximate. In the case of Mode 2, knowledge production is distributed, and the actors are far apart. The notion of distribution may be considered in five proximity dimensions (cognitive, organizational, social, institutional, geographical) [3].

Mode 2 of knowledge production has been increasingly applied in the research systems of many countries. This shift of science toward Mode 2 of knowledge production has occurred because Mode 2 is considered to be more interdisciplinary, more heterogeneous, closer to social actors and contexts, and more susceptible to social critique [4]. Mode 2 is an important factor in the increasing importance of indicators and indexes for assessment of research production of groups of researchers, since Mode 2 is connected to actions of teams consisting of several research groups.

In this chapter, additional indicators and indexes for assessment of research production of groups of researchers are discussed. From the viewpoint of bibliometric methodology, one may make a distinction among three levels of aggregations [5]: *micro level*—publication output of individuals and research groups; *meso level*—publication output of institutions or studies of scientific journals; *macro level*—publication output of regions and countries and groups of countries. We discuss below indexes belonging mainly to the meso level and the macro level of aggregation. These indexes may be applied to any organization that has components, and these components possess some units. Components may be researchers from a research group; research groups from a research institute or faculty; research institutes belonging to groups of institutes, etc. Units may be publications, citations, patents, etc. The following groups of indexes will be discussed:

1. *Simple indexes*: index of quality of scientific output; annual impact index; MAPR-index; T-index; RPG-index; TPP-index; TIA-index.
2. *Indexes for deviation from simple tendency*: Schutz coefficient of inequality; Wilcox deviation from the mode; Nagel's index of equality; coefficient of variation.
3. *Indexes for differences between components*: Gini's mean relative difference; Gini's coefficient of inequality.
4. *Indexes for concentration, dissimilarity, coherence, and diversity*: Herfindahl–Hirschmann index of concentration; Horwat's index of concentration; RTS-index of concentration; diversity index of Lieberson; generalized Stirling diversity index; index of dissimilarity; generalized coherence index.
5. *Indexes of imbalance and fragmentation*: Index of imbalance of Taagepera; RT-index of fragmentation.

6. *Indexes based on the concept of entropy*: Theil's index of entropy; redundancy index of Theil; negative entropy index; expected information content of Theil.
7. *Lorenz curve and associated indexes*; Lorenz curve, index of Kuznets; Pareto diagram.
8. *Indexes for the case of stratified data*: Index of Gini for stratified data; index of Kuznets for stratified data; coefficient of variation for stratified data; index of Theil for stratified data.
9. *Indexes of advantage and inequality*: Index of net difference of Lieberson; index of average relative advantage; index of inequity of Coulter; proportionality index of Nagel.
10. *RELEV method for assessment of scientific research*: Indexes and indicators connected to the RELEV method.
11. *Indexes and indicators for comparison among scientific communities in different countries*.
12. *Indexes and indicators for efficiency of research production from the point of view of publications and patents*.
13. *Indexes for characteristics of scientific production of a nation*.
14. *Indicators for leadership*.
15. *Selected journal citation measures*: Impact factor, intermediacy index; SNIP indicator; SJR.

Many examples for calculation of these indexes are provided. Special attention is devoted to calculation of the values of indexes for the two extreme cases (when one component possesses all the units and when all components possess the same number of units). Finally, we shall discuss the important question for research elites on the basis of a geometric detection of kinds of scientific elites from the Lorenz curve of the publications written by groups of researchers.

3.2 Simple Indexes

We shall discuss two indexes connected to citations of production of a group of researchers: the index for the quality of scientific output and the annual impact index. The remaining indexes discussed are connected with characteristics of the research publications of the group. They include the mean annual percentage rate (MAPR) index, the doubling-time index, the relative publication growth (RPG), and indexes of total publication productivity and total institutional authorship.

3.2.1 A Simple Index of Quality of Scientific Output Based on the Publications in Major Journals

Let us consider a hypothetical group of researchers (research group, department, institute, etc.). The group of researchers produces some output that is cited. Let us

count the citations for some time period (say one year or several years). One may consider the index

$$Q_1 = \frac{N_m}{N}, \tag{3.1}$$

where

- N: total number of citations of the research output of the group of scientists;
- N_m: number of citations of the research group in major journals.

In order to use this index, we need a list of major journals. If we have such a list for the corresponding scientific area, then the evaluators of scientific performance can use Q_1 as an orientation for the quality of the research output of the scientific group. In addition, some further analysis of N_m can be made. It may happen that:

1. Almost all of the N_m citations are citations of the output of a single person or of a small number of persons from the group of scientists. In this case, we have a group with one or several scientific leaders.
2. The citations N_m are more or less spread evenly among the scientific productions of all members of the group. In this case, we have a scientific group with some (smaller or larger) degree of homogeneity with respect to the quality of scientific output.

Let us discuss two examples of calculation of index of quality. We consider two research groups. Each group consists of five researchers. The first group consists of only young researchers. The number of citations (N_m, N) for the members of this group are (10, 15); (20, 31); (14, 22); (35, 48); (55, 62). Thus for the entire group, $N_m = 134$ and $N = 178$. Then $Q_1^I \approx 0.75$. The second group contains two established researchers. The number of citations for the members of this group are (753, 1042); (554, 782); (80, 119); (41, 56); (12,16). Thus for the entire group, $N_m = 1440$ and $N = 2011$. Then $Q_1^{II} \approx 0.72$, i.e., the quality index of the scientific output of the two groups is almost the same. This example was especially designed in order to show again that evaluation and comparison of research groups based on a single index is insufficient: in the one-dimensional space of the values of the simple index of quality, the two groups of researchers are very close one to each other. In order to evaluate them properly, we need a higher-dimensional space, i.e., we need sets of values of various indexes. These sets may represent the coordinates of the research groups in the multidimensional space of values of the indexes (quantitative evaluation space). A larger dimension of this space means more indexes to be used, and an increase in the dimension of the quantitative evaluation space usually increases the corresponding distance between points corresponding to the research groups in the space. The larger distance between research groups in the quantitative evaluation space allows better comparison of their research results.

3.2.2 Actual Use of Information Published Earlier: Annual Impact Index

The annual impact index for the ith year of the papers published in the nth year ($n < i$) $\mathrm{AI}_{i,n}$ [6] is defined as follows:

$$\mathrm{AI}_{i,n} = \frac{C_{i,n}}{P_n}, \tag{3.2}$$

where

- $C_{i,n}$: number of citations received in year i by the papers published in year n;
- P_n: number of papers, published in year n.

Let us fix n. When i is close to n, the annual impact index may increase with increasing i. Usually, at some value of i, the index has its maximum value, and when i increases further, the value of the index begins to decrease (one factor for this decrease is the aging of the information contained in the papers published in the year n).

The index of the actual use of information helps us to see easily whether the research information produced by a group of researchers is useful for the research society. Let us demonstrate this. We consider two research groups. Research group A has six publications for 2010, and research group B has twelve publications for 2010. The quantity of information produced by research group B is larger than that produced by A. The sets of citations of the above publications for the period 2011–2015 of the two groups are as follows:

- **Research group A**: 3, 8, 17, 38, 60;
- **Research group B**: 2, 8, 21, 49, 94.

The corresponding values of the $AI_{i,n}$-index are (approximately)

- **Research group A**: 0.5, 1.33, 2.83, 6.33, 12;
- **Research group B**: 0.16, 0.66, 1.75, 4.08, 7.83.

Thus according to the $AI_{i,n}$-index, the impact of the information produced by research group A is larger (at least for the five-year period of evaluation 2011–2015).

3.2.3 MAPR-Index, T-Index, and RPG-Index

These indexes are characteristics of the set of publications produced by the evaluated group of researchers [7, 8]. The MAPR-index (mean annual percentage rate) is defined as

$$\mathrm{MAPR}_t = 100\left[\frac{1}{t}\sum_{i=1}^{t}\frac{P_i - P_{i-1}}{P_{i-1}}\right], \tag{3.3}$$

where

- t: length of the studied period (in years);
- P_i: number of papers written by the group of researchers in year i

For example, if the period of evaluation is five years, then

$$\mathrm{MAPR}_5 = 20\left[\frac{P_1 - P_0}{P_0} + \frac{P_2 - P_1}{P_1} + \frac{P_3 - P_2}{P_2} + \frac{P_4 - P_3}{P_3} + \frac{P_5 - P_4}{P_4}\right]. \quad (3.4)$$

Note that all the P_i should be different from 0. The MAPR-index can also be used for characterization of the evolution of the number of publications in a research field or in a journal or group of journals.

The MAPR-index easily detects the phases of increasing or decreasing research activity. Let us consider two research groups that are evaluated for a period of five years ($t = 5$). Group A is a newly established research group, and group B is a mature group in a research field that is slowly beginning to be exhausted. The number of publications of the two groups are

- **Research group A**: 3, 5, 5, 7, 8, 11;
- **Research group B**: 63, 64, 62, 60, 58, 60.

The values of the MAPR_5-index for the two research groups are

- **Research group A**: $\mathrm{MAPR}_5^A \approx 1.583$;
- **Research group B**: $\mathrm{MAPR}_5^B \approx -0.346$.

The values of the MAPR-index that are very close to 0 or negative are evidence of maturity or of problems in the corresponding research group. The nature of such problems may be further studied by other quantitative or qualitative tools.

The T-index (the doubling time) is defined as

$$T = \frac{1}{2}\frac{0.301(t-1)}{\ln(P_t) - \ln(P_1)}, \quad (3.5)$$

where

- P_1: number of papers in the starting year;
- P_t: number of papers in the tth year.

The T-index gives a good impression about the mean size of the expansion of the scientific information produced by the research group or the mean size of expansion of information in a given research field. Let us consider two research fields with T-indexes of seven years and fifteen years. The first field expands faster. Faster expansion (and small value of the T-index) is characteristic for new fields or for established fields after a large discovery is made. The T-index of a mature field has a large value.

For the two research groups discussed above, we obtain the following values of the doubling-time index:

- **Research group A**: $T_A \approx 1.84$;
- **Research group B**: $T_B \approx -12.28$.

The results show that the tempo of advancing of the research activity of the newly established group is good, whereas the negative value of the index shows a shrinking of research production in the mature group B. Let us note that it is good practice to include only publications in journals of some level (i.e., journals with an impact factor or journals with an SJR factor) in order to achieve greater objectivity regarding the information supplied by the MAPR-index and by the T-index.

The RPG-index (relative publication growth index) [9] is defined as

$$\mathrm{RPG}_j(T) = \frac{P_j}{Q_j}; \quad Q_j = \sum_{i=1}^{T=j-1} P_i, \tag{3.6}$$

where

- $T = j - 1$: period in which the published papers are counted ($T \geq 3$);
- j: year for which the index is calculated;
- P_i: number of published papers in the ith year of the period of interest.

The value of T can be five years, ten years, twenty years, etc. The value of the RPG-index for several databases of papers (Chemical Abstracts, Biological Abstracts, Science Citations Index, etc.) can be found in Table 4.2 of [8]. The RPG index calculated with appropriately selected time periods may give us information about the dynamic equilibrium between recent information and previously published information.

As defined above, we can calculate, for example, $\mathrm{RPG}_{11}(10)$ but not $\mathrm{RPG}_{11}(8)$. In order to be able to calculate the value of the last index, we have to redefine the index slightly as follows:

$$\mathrm{RPG}_j(T = j - k) = \frac{P_j}{Q_j}; \quad Q_j = \sum_{i=k}^{j-1} P_i, \tag{3.7}$$

where $1 \leq k \leq j - 1$.

The $\mathrm{RPG}_5(4)$-index for the two research groups discussed in the subsection for the MAPR-index has the values

- **Research group A**: $\mathrm{RPG}_5^A(4) \approx 0.44$;
- **Research group B**: $\mathrm{RPG}_5^B(4) \approx 0.246$.

The result shows that the rate of total publications growth for research group A is about twice that of the rate for research group B.

3.2.4 Total Publication Productivity, Total Institutional Authorship

The TPP-index (total publication productivity index) [8] compares the total information productivity of groups of researchers working in fields with similar bibliometric features. The definition of the index is

$$\mathrm{TPP}_T = \frac{p_T}{\kappa T}, \tag{3.8}$$

where

- T: period of evaluation;
- κ: mean number of researchers working in the research group in the period T;
- p_T: total number of scientific publications published by the members of the research group in the period T.

As publications, one may count journal papers (also in electronic form), or in principle one may count any kind of scientific publications except conference abstracts.

The value of the TPP-index can be greatly influenced by multiple authorship. Because of this, it is useful if the TPP-index is accompanied by the TIA-index (total institutional authorship index) [10]

$$\mathrm{TIA}_T = \frac{A_a(T)}{A_t(T)}, \tag{3.9}$$

where

- T; period of evaluation;
- $A_a(T)$: Number of authors attributed to the evaluated research group for the period T;
- $A_t(T)$: total number of authors of the publications published by the research group for the period T.

3.3 Indexes for Deviation from a Single Tendency

The concept of indexes for deviation from a single tendency is as follows. One has a numerical series. By a mathematical operation one defines a value that is called the standard value (standard) for the series (different definitions can lead to different standards). Each value of the series deviates from the standard value. The indexes are constructed on the basis of these deviations.

In general, the tendency of deviation from a standard value can change over time. Below, we shall discuss mostly deviations from time-independent quantities. We just note that one can also construct deviations from time-dependent quantities. One such index is the Przeworski index of instability [11].

An important type of deviation from the time-independent quantities is deviations (absolute or squared) from a central tendency. Among the absolute deviations from a central tendency we shall discuss the indicators called the Schutz coefficient of inequality and the Wilcox deviation from the mode.

3.3.1 Schutz Coefficient of Inequality

The equation for this index is [12]

$$I_1 = \frac{\sum_{i=1}^{K} \left(\frac{P_i}{\overline{P}} - 1 \right)}{K - 1}, \tag{3.10}$$

where the quantities are as follows:

- K: number of components.
- P_i: percentage of the total number of units possessed by the ith component.
- $\overline{P}$: average percentage ($\overline{P} = (1/K) \sum_{i=1}^{K} P_i$).

Let us illustrate the extreme values of I_1 in terms of scientists and papers. If all scientists possess the same number of papers (absolute equality), then $P_i = \overline{P}$ and $I_1 = 0$. The other extreme case is that one of the scientists possesses all the papers and the other $K - 1$ scientists have none. Then the denominator of I_1 has the value $K - 1$, and $I_1 = 1$.

Inequality is an important concept in the social sciences and economics [13, 14]. Many measures developed for measuring economic and social inequality [15] can be used for measurement of different aspects of inequality of research production of researchers. Some of these indexes will be discussed below.

3.3.2 Wilcox Deviation from the Mode (from the Maximum Percentage)

The equation for this index is [16]

$$I_2 = 1 - \frac{\sum_{i=1}^{K} (P_m - P_i)}{K - 1}, \tag{3.11}$$

where the quantities are as follows:

- K: number of the components.
- P_i: percentage of the total number of units possessed by the ith component.
- P_m: the maximum value among $P_1, \ldots, P_k$.

I_2 measures the extent to which the nonmodal components resemble the modal component. In our example about scientists and papers, P_m is the share of the most productive scientist (measured by the number of published papers). If all the scientists are as productive as the most productive one, then $I_2 = 1$. If the most productive scientist wrote all the papers and the other scientists wrote none, then $I_2 = 0$, which indicates a problem.

The index of Wilcox was developed for measurement in political science. There exist several more indexes proposed by Wilcox [17] for measurement of different aspect of public opinion. These indexes (as has been shown above) can be easily adapted for assessment of research production.

Let us now turn to the squared deviations from a central tendency. Here we shall discuss Nagel's index of equality.

3.3.3 Nagel's Index of Equality

The equation for this index is [18]

$$I_3 = 1 - \frac{\sum_{i=1}^{K} \left(N_i - \frac{N}{K}\right)^2}{\left(Z - \frac{N}{K}\right)^2}. \tag{3.12}$$

The quantities above are as follows:

- N_i: Number of units possessed by the ith component of the organization;
- N: Total number of units distributed among the components;
- K: Number of components of the organization;
- Z: The worst possible allocation of components in terms of equality. This worst possible allocation occurs when one of the components owns all of the units and the other components own nothing.

Let the worst possible allocation be realized (a single researcher wrote all the publications in the research group, and the other researchers have none). Then $I_3 = 0$. And if all researchers from the research group wrote the same (N/K) number of publications, then $I_3 = 1$. Thus very small values of Nagel's index of inequality are evidence for the presence of a small number of highly productive researchers in the research group.

We note that the value of Nagel's index is sensitive to the number of components of the system.

3.3.4 Coefficient of Variation

The equation for the coefficient of variation is [19]

$$I_4 = \frac{1}{\overline{U}} \sqrt{\frac{1}{K} \sum_{i=1}^{K} (U_i - \overline{U})^2}, \tag{3.13}$$

where

- K: number of components of the organization;
- U_i: number of units owned by the ith component;
- $\overline{U}$: average number of units owned by the system components.

The variation coefficient is obtained by division of the standard deviation of the data by the mean value of the units owned by a component of the organization. Let the organization be a research group. One extreme case occurs when one component of the organization (one member of the research group) has written all the publications. Then $U_1 = U_2 = \cdots = U_{K-1} = 0$ and $U_K = K\overline{U}$. Then $I_4 = \sqrt{K-1}$. The other extreme case occurs when all researchers have written the same number of publications (namely $\overline{U}$ publications). Then $I_4 = 0$. Thus the presence of large differences in the research production of the researchers from the group will lead to a significant deviation of I_4 from 0. Another index of this kind is the logarithmic variance

$$I_5 = \frac{1}{K} \sum_{i=1}^{K} (\ln U_i - \ln \overline{U})^2. \tag{3.14}$$

If all researchers from the research group wrote the same number of publications, then $I_5 = 0$. In the extreme case that one of the researchers from the group wrote all the publications, then $I_5 = (\ln(K))^2/K$. We note that the indexes I_4 and I_5 can be easily normalized in order to have values between 0 and 1. We now proceed to the group of indexes for differences between components. Such indexes include, for example, the two quantities used by Gini: Gini's mean relative difference and Gini's coefficient of inequality.

3.4 Indexes for Differences Between Components

3.4.1 Gini's Mean Relative Difference

Gini's mean relative difference [20, 21] is calculated as follows:

$$I_6 = 1 - \frac{\sum_{i=1}^{K-1} \sum_{j=i+1}^{K} \mid P_i - P_j \mid}{K-1}, \tag{3.15}$$

where the quantities are

- K: number of components of the organization;
- P_i: percentage of the total number of units possessed by the ith component.

The values of I_6 are between 0 and 1. When one of the components possesses all units, the value of I_6 is 0, regardless of the number of components. When all components possess the same number of units, the value of I_6 is 1. Let the units be the publications written by the researchers from a research group. If one of the researchers wrote all the publications, then Gini's mean relative difference is 0. If all researchers wrote the same number of publications, then the value of the index is 1.

Gini's mean relative difference also has a continuous version [22], which was used for quantification of the speed of technological adoption in India. An extensive discussion on the measures of Gini and similar measures such as the Lorenz curve can be found in [23, 24].

We note that the values of I_6 do not correspond to expectations that might arise from the name of the index. One might expect that the value 0 will be assigned to the case in which no difference between researchers exists (all of them wrote the same number of publications). And for the extreme case (one researcher wrote all publications), the expectation for the value of the index is that it should be equal to 1. The real situation is exactly the opposite, and this is one factor that contributes to the popularity of the following index: I_7.

3.4.2 Gini's Coefficient of Inequality

Gini [20] preferred to use I_6, but in the course of time, Gini's coefficient of inequality

$$I_7 = \frac{1}{2} \sum_{i=1}^{K} \sum_{j=1}^{K} \left| \frac{P_i - P_j}{K} \right| \tag{3.16}$$

become more popular. The quantities in I_7 are as follows:

- K: number of components of the organization;
- P_i: percentage of the total number of units possessed by the ith component.

Gini's coefficient is sensitive to the number of components K, and because of this, it is better when I_7 is used in organizations that have a large number of components K. If the number of components is small, then it is better to use I_6.

Let us calculate Gini's coefficient of inequality for several cases of groups of researchers and their research publications. If all researchers wrote the same number of publications, then $I_7 = 0$. If one of researchers wrote all publications, and other researchers wrote none, then $I_7 = K - 1$. (Thus the index can be normalized when one divides it by $K - 1$: $I_7^+ = I_7/(K - 1)$.) Let us now suppose we have a research group of five researchers and the percentage of publications they wrote is $P_1 = 0.15$;

$P_2 = 0.18$, $P_3 = 0.22$, $P_4 = 0.30$, $P_5 = 0.15$. Then the value of Gini's coefficient of inequality is $I_7 = 0.101$. Thus the value of I_7 is closer to 0 than to the maximum value of 4, which reflects the fact that the inequality with respect to the number of publications in the group of researchers is not very large.

Gini's coefficient is much used in economics, the social sciences, ecology, etc. [25–30]. An example of its use in the area of scientific research is for quantification of the concentration of scientific research and innovation [31].

3.5 Indexes of Concentration, Dissimilarity, Coherence, and Diversity

The next group of indexes are indexes for concentration and diversity. These indexes inform us how the quantities associated with research production (number of publications, number of citations, etc.) are concentrated among groups of researchers. An exploration of the concentration of research production reveals also fragmentation, diversity, coherence, and imbalance with respect to scientific production in research organizations. Diversity may be defined as the property of apportioning units into categories in any system [32]. Coherence may be defined as the property of relating categories via units. Coherence captures the extent to which the various parts in a system are directly connected via some relation. Diversity has the following three distinct attributes: (i) *variety*—number of categories into which the units are apportioned; (ii) *balance*—evenness of the distribution of units across categories; (iii) *disparity*—degree to which the categories of the units are different [4]. The diversity of a system increases not only with more categories (higher variety) and with a more balanced distribution (higher balance), but also if the units are allocated to more different categories (higher disparity). Coherence has the following attributes: (i) *density*—number of relations between categories; *intensity*—overall intensity of the relations in the system; (iii) *disparity*—degree to which the categories of the relations are different.

In the process of analysis of diversity, one may use units such as university, institute, faculty, department, article, researcher, and research topic such as an emergent technology. Some of these units may be connected to a small number of the corresponding items. Thus one may not have enough items for a robust statistical analysis, which may worsen the quality of the resulting measures.

3.5.1 Herfindahl–Hirschmann Index of Concentration

The equation for this index [33, 34] is

$$I_8 = \sum_{i=1}^{K} P_i^2, \tag{3.17}$$

where

- K: number of components of the organization;
- P_i: percentage of the total number of units possessed by the ith component.

This form of the index is insensitive to small values of P_i, since the square of a value that is close to 0 is quite a small number. The index I_8 has its maximum value of 1 when one of the components of the organization possesses all units (in the case of our example, when one of the scientists possesses all the papers). The minimum value of the index is $1/K$ when all the components possess an equal number of units (there is no concentration of papers). Thus the lower bound of the index depends on the number of components K. In order to avoid this and to bound I_8 between 0 and 1, one can use the following form of the index:

$$I_8^* = 1 - \frac{1 - I_8}{(1 - 1/K)}. \tag{3.18}$$

When the number of components (the number of researchers) is large, then $1/K$ is small, and one can use I_8. If, however, the number of components is small, then it is better to use I_8^*.

Let us calculate I_8 for the case of the group discussed above for the case of index I_7. The result is $I_8 = 0.2158$, which reflects the relatively small level of concentration of ownership of research publications in the evaluated research group.

The Herfindahl–Hirschmann index has been used for measurement of dominant power [35], for measuring concentration in portfolio management [36], etc. [37, 38].

3.5.2 *Horvath's Index of Concentration*

The equation for this index is [39]

$$I_9 = P_m + \sum_{i=2}^{K} P_i^2[1 + (1 - P_i)], \tag{3.19}$$

where

- K: number of components of the organization;
- P_i: percentage of the total number of units possessed by the ith component;
- P_m: percentage of the total number of units possessed by the modal component (the component that possesses the largest number of units).

Horvath's concentration index measures the influence of the largest component. In our example, the modal component consists of the researcher with the largest number of publications. The index is useful in cases in which one of the scientists dominates the group of scientists with respect to some quantity (for example the number of published papers). The index I_9 measures the change in the primacy of this researcher

within the group in the course of time. Let us illustrate this. We shall consider a research group of five researchers. At the beginning, one of the researchers possesses all the publications of the group, and the other (young) researchers have not written any publications. In this case, $I_9 = 1$. In two years, the situation changes. The experienced researcher still dominates with 90 % of the papers, but the other four researchers have also written some papers. Let the percentage distribution be 0.9, 0.04; 0.02; 0.02; 0.02. Then the value of the index is $I_9 = 0.95512$, which reflects the changes but still shows the dominance of the most experienced researcher from the evaluated research group.

3.5.3 RTS-Index of Concentration

This index was designed by Ray et al. [40, 41]. The equation for this index is

$$I_{10} = \left[\frac{\sum_{i=1}^{K} P_i^{\alpha} - K^{(1-\alpha)}}{1 - K^{(1-\alpha)}} \right]^{(1/\alpha)}, \tag{3.20}$$

where

- K: number of components of the organization;
- P_i: percentage of the total number of units possessed by the ith component;
- α: parameter.

A characteristic feature of this index is that it depends on the parameter α. For $\alpha = 0$, $I_{10} = 0$. For $\alpha = 1$, $I_{10} = 1$. As $\alpha \to \infty$, $I_{10} \to P_m$, where P_m is the modal share of units (the number of units of the largest possessor of units).

Indexes of concentration are quite useful in the evaluation of research groups. They can exhibit hidden problems, such as concentration of research publications in researchers who are at the end of their scientific career, which hints at a future decrease in research productivity of this research group.

3.5.4 Diversity Index of Lieberson

The equation for this index is [42]

$$I_{11} = \frac{1 - \sum_{i=1}^{K} P_i^2}{(1 - 1/K)}, \tag{3.21}$$

where

- K: number of components of the organization;
- P_i: percentage of the total number of units possessed by the ith component.

The index I_{11} is bounded between 0 and 1. Let us discuss a group of researchers and their research publications. If one of the researchers owns all publications, then $I_{11} = 0$, and if all researchers have written the same number of publications, then $I_{11} = 1$. As an example for application of the index of diversity, let us consider two research groups. Research group A consists of five researchers, and the percentages of research publications are as follows 0.3, 0.25, 0.2, 0.15, 0.1. Research group B consists of six researchers, and the percentages of research publications are 0.25, 0.2, 0.15, 0.15, 0.15, 0.1. The values of the index are as follows:

- **Research group A**: $I_{11}^{A} = 0.96875$;
- **Research group B**: $I_{11}^{B} = 0.984$.

Thus the diversity of the two research groups is almost the same, and the value of the index is close to 1, which hints at sufficient activity of all researchers from the evaluated research groups.

3.5.5 Second Index of Diversity of Lieberson

Let us consider two populations Q and R. Now we want to study the diversity between the populations with respect to some category. The equation for the index is [42]

$$I_{12} = 1 - \sum_{i=1}^{C} Q_i R_i, \tag{3.22}$$

where

- Q_i: proportion of the category in population Q;
- R_i: proportion of the category in population R;
- C: the number of categories.

The populations Q and R can be of any type. For example, they may be the populations of researchers in two research institutes. The category can be any nominal category of some attribute. For example, the attribute can be the age of researchers and the categories can be young researchers (up to age 40); intermediate-age researchers (40–60 years old), and mature researchers (over 60 years old).

The index I_{12} reaches its maximum value of 1 when the diversity between the two populations is maximal. This happens when, for example, all Q_i equal 0 and all R_0 are positive.

Let us consider one example. We have two research institutes from the same area (say physics). For institute A, the percentage of young researchers is 0.05, the

percentage of intermediate age researchers is 0.15, and the percentage of mature researchers is 0.8. In = institute B, the percentage of young researchers is 0.08, the percentage of intermediate-age researchers is 0.25, and the percentage of mature researchers is 0.67. The index of diversity of Lieberson for these two institutes is $I_{12} = 0.4325$.

The diversity index of Lieberson can be used for analysis of different kinds of networks [43], electoral competition [44], etc.

3.5.6 Generalized Stirling Diversity Index

Let us consider units of something (e.g., publications) distributed among N categories (e.g., categories connected to the ISI Web of Science). Let p_i be the proportion of the units in category i, and d_{ij} the distance between categories i and j. Then the generalized Stirling diversity index is [32]

$$S = \sum_{i,j(i\neq j)} (p_i p_j)^{\alpha} d_{ij}^{\beta}, \tag{3.23}$$

where α and β are parameters. In order to use this index, one has to choose appropriate categories and to assign units to each category. Then one has to construct adequate metrics for the distance d_{ij} and to set appropriate values of the parameters α and β. Often one chooses the density in the interval $0 < d_{ij} < 1$, and the choice of small values of β emphasizes the importance of distance for the studied problem.

Particular cases of the generalized Stirling diversity index are the Rao–Stirling diversity index ($\alpha = \beta = 1$) [45, 46]

$$S_{RS} = \sum_{i,j(i\neq j)} (p_i p_j) d_{ij}; \tag{3.24}$$

and the Simpson diversity index ($\alpha = 1; \beta = 0$)

$$S_S = \sum_{i,j(i\neq j)} (p_i p_j) = 1 - \sum_i p_i^2. \tag{3.25}$$

The Rao–Stirling index may be interpreted as the average cognitive distance between elements, as seen from the categorization, since it weights the cognitive distance d_{ij} over the distribution of elements across categories [4]. The Rao–Stirling diversity index can be added over scales (under some plausible assumptions) [47]. Then, for example, the diversity of a research institute is the sum of the diversities within each article it has published, plus the diversity between the articles. This interesting property leads to the possibility of measuring the diversity of large organizations in a modular manner.

3.5.7 Index of Dissimilarity

Let us have two groups of researchers that are classified with respect to some characteristic that has two possible values (for example, one group consists of researchers who have published papers, and the second group consists of researchers who have not published even a single paper). The equation for the index is

$$I_{13} = \frac{1}{2} \sum_{i=1}^{K} \mid G_{1i} - G_{2i} \mid, \tag{3.26}$$

where

- K: number of investigated research organizations;
- G_{1i}: proportion of components of the ith organization that can be characterized by the first value of the characteristics;
- G_{2i}: proportion of components of the ith organization that can be characterized by the second value of the characteristics.

Let us now consider two research groups. Research group A has ten members, and eight of them have publications. Research group B has fourteen members, and eleven of them have publications. In this case, $I_{13} = 0.015$. Let now two new PhD students join research group B. Thus it has sixteen members, and eleven of them have publications. The value of the index changes to $I_{13} = 0.1175$, which reflects the fact of increasing dissimilarity and diversity between the two groups of researchers.

In its original definition [48], I_{13} was defined as an index of segregation (for example, segregation of citizens of different skin color in some urban area).

3.5.8 Generalized Coherence Index

Let us consider units of something (e.g., publications) distributed among N categories (e.g., categories connected to the ISI Web of Science). Let p_i be the proportion of units in category i; I_{ij} the intensity of relations between categories i and j; and d_{ij} the distance between categories i and j. Let us suppose that we have constructed adequate metrics for distance and intensity. The generalized coherence index [4] is given by the equation

$$G = \sum_{ij(i \neq j)} I_{ij}^{\gamma} d_{ij}^{\beta}. \tag{3.27}$$

When $\gamma = \delta = 0$, the value of G is equal to M. For $\gamma = 1$ and $\delta = 0$, we obtain a measure of intensity

$$G_I = \sum_{ij(i \neq j)} I_{ij} = 1 - \sum_{i} I_{ii}, \tag{3.28}$$

and for $\gamma = \delta = 1$, we obtain a measure of coherence

$$G = \sum_{ij(i \neq j)} I_{ij} d_{ij}. \tag{3.29}$$

If the intensity of relations is defined as the distribution of relations (i.e., when I_{ik} is equal to p_{ik}), then the coherence from (3.29) may be interpreted as the average distance over the distribution of relations p_{ik}.

3.6 Indexes of Imbalance and Fragmentation

The next group of indexes consists of indexes of imbalance and fragmentation. From among these indexes, we shall discuss the index of imbalance of Taagepera and the RT-index of fragmentation.

3.6.1 Index of Imbalance of Taagepera

This index treats imbalance as a comparison of the size of the largest component with respect to the size of the next-largest one. The equation for the index is [49]

$$I_{14} = \frac{\sum_{i=1}^{K-1} \frac{(P_i - P_{i+1})}{i} - (\sum_{i=1}^{K} P_i^2)^2}{\sqrt{\sum_{i=1}^{K} P_i^2 - (\sum_{i=1}^{K} P_i^2)^2}}, \tag{3.30}$$

where the components of the organization are ranked in decreasing order with respect to the possessed units and

- K: number of components of the organization;
- P_i: percentage of the total number of units possessed by the ith component.

The index I_{14} is most sensitive to the size difference (called imbalance) between the two largest components of the organization. A larger difference leads to a larger value of I_{14}.

3.6.2 RT-Index of Fragmentation

The relationship for this index is [50]

$$I_{15} = 1 - \frac{\sum_{i=1}^{K} N_i(N_i - 1)}{N(N-1)}, \tag{3.31}$$

where

- K: number of components of the organization;
- N_i: total number of units possessed by the ith component;
- N: total number of units possessed by all components of the organization.

The index is designed as 1 minus a measure of concentration of units among the components of the organization. In our example, the concentration of all papers to the account of one scientist leads to $I_{15} = 0$. When the papers are uniformly distributed among the scientists, then I_{15} is roughly equal to $1 - 1/K^2$, and for a large number of components of the organization, this value is almost equal to 1. From the last sentences, it follows that one has to use I_{15} for evaluation of fragmentation in organizations that have a large enough number of components.

We stress the following characteristic of I_{15}. If two groups of researchers (each with some fragmentation with respect to the possession of their published papers) are combined into a single group, then I_{15} for the new group will have a larger value than the values for the two groups considered separately. In other words, when groups are combined, then I_{15} shows a greater fragmentation in the new group in comparison to the two groups that are combined.

3.7 Indexes Based on the Concept of Entropy

Most of the indexes discussed below have the useful properties of **aggregation** and **decomposition**. The decomposition property means that the corresponding measure (of inequality in research productivity, for example) for the entire population of researchers (of a research group, research institute, etc.) can be decomposed as a sum of measures within the subpopulations (within the sections of the institute). Aggregation means the opposite: the sum of the corresponding measures for the subpopulation gives the value of the measure for the entire population.

The concept of entropy is used in analyses of science dynamics [51]. In order to understand the indexes based on the concept of entropy, we need the following concepts:

- **Bit**: Let us have m alternatives and we have to choose one of them. The number of bits of information h needed to select one of these alternatives is defined as $m = 2^h$. Then $h = \log_2 m$. In other words, one bit of information is gained when the value of a specific random variable (a variable that can take the value 0 or 1 with equal probability) becomes known.
- **Entropy of a set of random variables**: Let us have a set of L random variables each of which has its own probability of occurrence p_i and its own information

of h_i bits. The entropy of the set equals the sum of the information values of all the individual variables, each weighted by the corresponding probability of occurrence:

$$H = \sum_{i=1}^{L} p_i h_i = \sum_{i=1}^{L} p_i \log_2(1/p_i) = -\sum_{i=1}^{L} p_i \log_2(p_i).$$

The maximum value of the entropy is obtained when all probabilities of occurrence are the same. When one of the probabilities of occurrence is close to 1 (and the others are close to 0), then H is close to 0.

3.7.1 Theil's Index of Entropy

The probabilities p_i discussed above can be interpreted as percentages of the total number of units possessed by the ith component. In such a way, the entropy can be used directly as a measurement of (scientific) inequality. The result is Theil's index of entropy. The equation for the index is [52–54]

$$I_{16} = -\sum_{i=1}^{K} P_i \log_2 P_i, \tag{3.32}$$

where

- K: number of components of the organization;
- P_i: percentage of the total number of units possessed by the ith component.

A larger value of I_{16} corresponds to greater equality in the group of components (which means that the differences among the numbers of published papers among the scientists from the studied group is not very large).

Let us calculate I_{16} for several cases for a group of researchers and their research publications. Let one of researchers own all of publications, and the other members of groups have written no publications. There will be a difficulty in calculating I_{16} if some of the researchers have no publications, but we can assume that the contribution of the corresponding term to the index is 0. Then $I_{16} = 0$. For the case that all researchers have written the same number of publications, the value of the index is $I_{16} = \log_2 K$. The last result shows that I_{16} can be rescaled as follows:

$$I_{16}^{*} = -\frac{\sum_{i=1}^{K} P_i \log_2 P_i}{\log_2 K}. \tag{3.33}$$

Let us suppose a group of four researchers and that the percentages of publications that they have written are 0.5, 0.3, 0.1, 0.1. Let us have another group of eight

researchers with percentages of publications 0.3, 0.15.0.15.0.15, 0.1, 0.1, 0.03, 0.02. The values of Theil's index of entropy are

- **Research group A**: $I_{16}^{*\,A} \approx 0.84$;
- **Research group B**: $I_{16}^{*\,B} \approx 0.89$,

which means that the level of equality in group B with respect to research publications is slightly greater than the equality in research group A.

Theil's index is much used in sociology [55] and in economics [56].

3.7.2 Redundancy Index of Theil

The equation for this index is [57, 58]

$$I_{17} = \log_2 K + \sum_{i=1}^{K} P_i \log_2 P_i, \tag{3.34}$$

where

- K: number of components of the organization
- P_i: percentage of the total number of units possessed by the ith component.

The index I_{17} is an index of concentration, since we subtract the absolute entropy from a certain constant value. This index can be normalized as follows:

$$I_{17}^{*} = \frac{\log_2 K + \sum_{i=1}^{K} P_i \log_2 P_i}{\log_2 K}. \tag{3.35}$$

For the two research groups studied by means of I_{17}^{*}, one obtains the following values of the normalized redundancy index of Theil:

- **Research group A**: $I_{17}^{*\,A} \approx 0.16$;
- **Research group B**: $I_{17}^{*\,B} \approx 0.11$,

which shows that the concentration of publications in research group A is greater than that of research group B.

3.7.3 Negative Entropy Index

The equation for this index is

$$I_{18} = \text{antilog}_2\left(-\sum_{i=1}^{K} P_i \log_2 P_i\right), \tag{3.36}$$

where

- K: number of components of the organization;
- P_i: percentage of the total number of units possessed by the ith component.

The antilog function is the inverse of the log function. In (3.36), we use 2 as the base of the log and antilog functions. In the original definition of the index [59], the base was 10.

In our examples about researchers and their publications, I_{18} measures the closeness in the values of the numbers of publications written by every researcher. The index can be normalized as follows:

$$I_{18}^{*} = \frac{\text{antilog}_2\left(-\sum_{i=1}^{K} P_i \log_2 P_i\right)}{K}. \tag{3.37}$$

3.7.4 Expected Information Content of Theil

Let us suppose that we have a message that *an a priori distribution* $\sum p_i$ *has turned into an a posteriori distribution* $\sum q_i$. The expected information content of this message is [60]

$$I = \sum_i q_i^2 \log \frac{q_i}{p_i}. \tag{3.38}$$

If the logarithm has base of 2, then I is expressed as bits of information. Leydesdorff [51] has used this index in order to study statistics of journals from the SCI Journal Citation Reports.

3.8 The Lorenz Curve and Associated Indexes

3.8.1 Lorenz Curve

In general, the Lorenz curve can be defined as follows [61, 62]. Let us assume a probability distribution $P = F(x)$ of some quantity (number of papers, number of citations, amount of money, etc.) owned by members of some class of people (such as scientists) and let x be normalized in such a way that its value is between 0 and 1. The inverse distribution of F is $x = F^{-1}(P)$, and the Lorenz curve is defined by

$$L(F) = \int_0^1 F^{-1}(P)dP. \tag{3.39}$$

Let us assume a group of K researchers, and suppose we are interested in constructing the Lorenz curve for the number of papers written by every scientist. Let us rank the scientists with respect to the number of papers written by them. Let n_i be the number of papers of the ith scientist from the ranked list (the ranking is made in such a way that $n_1 \leq n_2 \leq \ldots n_K$). Then the coordinates of the corresponding Lorenz curve are

$$F_i = \frac{i}{K}; \quad L_i = \frac{\sum_{j=1}^{i} n_i}{\sum_{i=1}^{K} n_i}. \tag{3.40}$$

The Lorenz curve is much used in research on income distributions [63, 64], land use [65], economic concentration [66], etc. [67]. The Lorenz curve is used in scientometrics for characterization of conjugate partitions [68], for measurement of relative concentration [69, 70], group preferences [71], distribution of publications [72], distribution of research grants [73], regional research evaluation [74], and university ranking [75].

3.8.2 *The Index of Gini from the Point of View of the Lorenz Curve*

The points (0, 0); (0, 1); (1, 0); (1, 1) determine a square in the (L, F)-plane. The diagonal of this square that connects (0, 0) and (1, 1) is called the line of absolute equality: all components of the organization possess the same number of units. In practice, there is no absolute equality, and in this case, the Lorenz curve is below the line of absolute equality. Then a region exists between the line of absolute equality and the Lorenz curve. The area of this region is connected to the index of Gini:

$$I_{19}^{\dagger} = 1 - 2\int_0^1 L(F)dF. \tag{3.41}$$

The discrete version of the index of Gini is closely connected to the Gini coefficient of inequality (I_7) discussed above. The difference is that the index of Gini is divided also by the mean number of units $\overline{U}$ owned by a system component:

$$I_{19} = \frac{1}{2K^2\overline{U}} \sum_{i=1}^{K} \sum_{j=1}^{k} \mid U_i - U_j \mid, \tag{3.42}$$

where

- K: number of components of the organization;
- U_i: number of units owned by the ith component;
- $\overline{U}$: average number of units owned by the system components.

If the components are ranked with respect to the units they own ($U_1 \geq U_2 \geq \cdots \geq U_K$), then the equation for the index of Gini is

$$I_{19} = 1 + \frac{1}{K} - \frac{2}{K^2 \overline{U}} \sum_{i=1}^{K} i U_i. \tag{3.43}$$

3.8.3 Index of Kuznets

The equation for this index is [19]

$$I_{20} = \frac{1}{2K\overline{U}} \sum_{i=1}^{K} \mid U_i - \overline{U} \mid . \tag{3.44}$$

where

- K: number of components of the organization;
- U_i: number of units owned by the ith component;
- $\overline{U}$: average number of units owned by the system components.

The index of Kuznets has a form that is similar to that of the index of Gini, discussed above. There is, however, a difference. In the case of the index of Gini, one compares each component to each other component with respect to the number of possessed units (papers, citations, or money). In the case of the index of Kuznets, the comparison is different: the number of units possessed by each component is compared to the mean number of possessed units.

3.8.4 Pareto Diagram (Pareto Chart)

The Pareto diagram, also called a Pareto chart, is famous in the area of econometrics [76, 77]. In general, it is constructed as follows. On the abscissa of the coordinate system one puts the logarithm of the number of units (number of citations, for example). On the ordinate, one puts the logarithm of the relative cumulative frequencies (of the number of scientists that have the corresponding number of citations).

It can happen (as happens often in econometrics) that some of the points are approximately on a straight line (Pareto line). Then the angle between the Pareto line and the abscissa (the coefficient α of Pareto) is a characteristic measure of the corresponding distribution.

3.9 Indexes for the Case of Stratified Data

In some cases, the empirical data are stratified into layers. For example, we know the number of researchers who have published between zero and five papers; then the number of researchers who have published between six and ten papers, etc. We do not know the distribution within the layers (e.g., we do not know how many scientists have written seven papers). In addition, it may happen that the sizes of the different layers are not the same.

There are equations for many of the indexes for the case of stratified data. For the indexes discussed above, some of the equations are as follows [19]:

- Index of Gini for stratified data. The equation is

$$I_{19}^{*} = \left[\sum_{i=1}^{M} \left(2 \sum_{j=1}^{i} P_i - P_j \right) P_i \frac{U_i}{\overline{U}}, \right] - 1 \tag{3.45}$$

  where

  – M: number of layers for the stratified data;
  – P_i: percentage of the total number of units possessed by the ith component;
  – U_i: number of units owned by the ith component;
  – $\overline{U}$: average number of units owned by the system components,

  where U_i are ordered as follows: $U_1 \leq U_2 \leq \cdots \leq U_M$.
- Index of Kuznets for stratified data. The equation is

$$I_{20}^{*} = \frac{1}{2} \sum_{i=1}^{M} P_i \mid \frac{U_i}{\overline{U}} - 1 \mid, \tag{3.46}$$

  where

  – M: number of layers for the stratified data;
  – P_i: percentage of the total number of units possessed by the ith component;
  – U_i: number of units owned by the ith component;
  – $\overline{U}$: average number of units owned by the system components.
- Coefficient of variation for stratified data. The equation is

$$I_{4}^{*} = \frac{1}{\overline{U}} \sqrt{\sum_{i=1}^{M} (U_i - \overline{U})^2 P_i}, \tag{3.47}$$

  where

  – M: number of layers for the stratified data;
  – P_i: percentage of the total number of units possessed by the ith component;

– U_i: number of units owned by the ith component;
– $\overline{U}$: average number of units owned by the system components.

The equation for the coefficient of logarithmic variance is

$$I_5^* = \sum_{i=1}^{M} \left(\ln \frac{U_i}{\overline{U}} \right)^2 P_i. \tag{3.48}$$

- Index of Theil. The equation is

$$I_{21} = \sum_{i=1}^{M} P_i \frac{U_i}{\overline{U}} \log_2 \left(\frac{U_i}{\overline{U}}, \right) \tag{3.49}$$

where

– M: number of layers for the stratified data;
– P_i: percentage of the total number of units possessed by the ith component;
– U_i: number of units owned by the ith component;
– $\overline{U}$: average number of units owned by the system components.

Up to now, we have discussed a group of researchers. When one has to compare several groups of researchers (for example, several institutes of physics belonging to a national research institution), one may use additional indexes. Some of them will be discussed below.

3.10 Indexes of Inequality and Advantage

3.10.1 Index of Net Difference of Lieberson

The equation for this index is [79]

$$I_{22} = \sum_{i=1}^{I} A_i \left(\sum_{j=1}^{i-1} B_j \right) - \sum_{i=1}^{I} B_i \left(\sum_{j=1}^{i-1} A_j \right), \tag{3.50}$$

where

- I: number of classes;
- i: a class of the ranked distribution of the classes;
- A_i: proportion of units of group A in the class i;
- B_i: proportion of units of group B in the class i;
- $\left(\sum_{j=1}^{i-1} A_j \right)$: cumulative percentage of units of group A ranked below class i;

- $\left(\sum_{j=1}^{i-1} B_j\right)$: cumulative percentage of units of group B ranked below class i.

Within the scope of our example about the researchers and their publications, the application of the index can be as follows, for example. Let us define $I = 6$ classes: between 0 and 10 papers; between 11 and 20 papers; between 21 and 30 papers; between 31 and 40 papers; between 41 and 50 papers; and over 50 papers. Let us define the two groups of researchers as follows:

- group A: young researchers up to 40 years old;
- group B: researchers over 40 years old.

Then I_{22} will measure the net difference between the young and mature researchers with respect to the six classes defined above (and connected to the number of papers written by a scientist).

The index of net difference of Lieberson can be used to investigate segregation [80]. In the area of scientific systems and structures, the index has been used, for example, for studying the distribution of scientific positions for women in Israel [81].

3.10.2 Index of Average Relative Advantage

The equation for this index is [82]

$$I_{23} = \sum_{i=1}^{I} \sum_{j=1}^{J} k_{ij} A_i B_j, \tag{3.51}$$

where

- A_i: proportion of units of group A in the class i;
- B_j: proportion of units of group B in the class j;

and k_{ij} is a coefficient that has values as follows:

- $k_{ij} = \frac{A_i - B_i}{A_i}$ if $A_i > B_i$;
- $k_{ij} = 0$ if $A_i = B_i$, $k_{ij} = \frac{A_i - B_i}{B_i}$ if $A_i < B_i$.

This index accounts for all possible pairwise combinations, and it weights them by a coefficient that is proportional to the relative magnitude of the advantage involved (where the advantage is understood as a larger share of the units of class A in comparison to the units of class B).

The index of average relative advantage has been used to study the advantages and disadvantages of social groups with respect to jobs, income, education, etc. But this index can also be used to study groups of researchers with respect to the characteristics of their scientific production (such as number of papers or number of citations).

Let us now consider two indexes of inequity. These indexes measure the deviation from uniformity in some distribution.

3.10.3 Index of Inequity of Coulter

The equation for this index is [83]

$$I_{24,\alpha} = \frac{[\sum_{i=1}^{K} \mid P_i - Q_i \mid^{\alpha}]^{(1/\alpha)}}{[(1 - \min(Q))^{\alpha} - (\min Q)^{\alpha} + \sum_{k=1}^{K} Q_i^{\alpha}]^{(1/\alpha)}}, \tag{3.52}$$

where

- P_i: the proportional share of a component;
- Q_i: the proportional share that should be received by the component with respect to the equity standard distribution;
- $\min(Q)$: the smallest value of Q;
- α: a value that is set by the investigator. The value of α determines the sensitivity of the index to concentration. Thus an appropriate choice of α makes the index sensitive not only to inequality but also to concentration to the degree that is desired by the investigator.

The inequality index of Coulter may be used in the analysis of possible locations of different facilities (including scientific facilities) [84].

3.10.4 Proportionality Index of Nagel

The equation for this index is [18]

$$I_{25} = 1 - \frac{\sum_{i=1}^{K}(P_i - A_i)^2}{\sum_{i=1}^{K}(Q_i - A_i)^2}, \tag{3.53}$$

where

- P_i: actual frequency distribution of the units to the components (proportion of units assigned to the ith component);
- A_i: distribution of units to the components in proportion to merit (standard distribution—shares that would occur if the units were distributed in proportion to an equity standard such as merit);
- Q_i: zero allocation (the most inequitable distribution of units to the components possible). Often the distribution treats the case in which one of the components owns all the units and the other components do not possess anything.

In our example, P_i is proportional to the number of publications of the ith researcher; A_i reflects the situation in which all researchers have the same number of publications. And the values of Q_i correspond to the situation that one of the researchers has written all the publications and the other researchers have written none.

The frequency of quantitative evaluations of national research systems has been increasing [85–93]. Because of this, we shall discuss below the following methods and sets of indicators and indexes for performing such evaluation: the RELEV method for assessment of scientific research performance within public institutes; indexes and indicators for comparison among scientific communities in different countries; efficiency of research production from the point of view of publications and patents, etc.

3.11 The RELEV Method for Assessment of Scientific Research Performance in Public Institutes

The RELEV method [94–97] assigns a single numerical value to the research performance of a research institute. With respect to this value, different institutes working on closely related research fields can be compared. The index provided by the RELEV method can be a useful addition to the basket of indexes that form the quantitative part of research evaluation in a system of research institutions. The definition of the index for the ith institution from the set of compared institutions is as follows:

$$\Omega_{\mathrm{RELEV}}(i) = 3 - X_{1i} + X_{2i} + X_{3i} + X_{4i} + X_{5i} + 2X_{6i} + X_{7i}, \tag{3.54}$$

where seven indexes connected to the evaluated n institutions are taken into account:

1. A: Index of public funds attributed to the research institutions, $(\alpha_1, \ldots, \alpha_n)$;
2. B: index of self-financing (funds attracted by the research institution in addition to the public funds), $(\beta_1, \ldots, \beta_n)$;
3. X: index of personnel in training (number of trained individuals), $\xi_1, \ldots, \xi_n$;
4. Δ: index of teaching activities of researchers (hours of teaching by the scientists), $\delta_1, \ldots, \delta_n$;
5. E: index of national publications (numbers of national publications), $\varepsilon_1, \ldots, \varepsilon_n$;
6. Φ: index of international publications (number of international publications), $\phi_1, \ldots, phi_n$;
7. Γ: patent index (number of patents), $\gamma_1, \ldots, \gamma_n$.

The indexes above can be calculated in two ways: per researcher; as the total number for the corresponding institution. Our experience shows that in most cases, it is more reasonable to calculate the above indexes per researcher.

Let $max_A, max_B, max_X, max_\Delta, max_E, max_\Phi, max_\Gamma$ be the maximum values of corresponding indexes in the set of evaluated institutions. Then

$$X_{1i} = \alpha_i / max_A;\ X_{2i} = \beta_i / max_B;\ X_{3i} = \xi_i / max_X;\ X_{4i} = \delta_i / max_\Delta;$$
$$X_{5i} = \varepsilon_i / max_E;\ X_{6i} = \phi_i / max_\Phi;\ X_{7i} = \gamma_i / max_\Gamma. \quad (3.55)$$

Some of the indexes can have larger weights, as, for example, the index X_{6i} connected to publications of international journals. Weight coefficients can be introduced for all indexes, and this is a main direction of work on adjustment of the RELEV method in evaluating institutions [95, 97]. In addition, the number of indexes can be increased or some of the indexes can be changed. This depends on the specifics of the evaluated institutions.

3.12 Comparison Among Scientific Communities in Different Countries

Countries can be compared with respect to different characteristics of their scientific communities. For this, one needs an appropriate system of indicators and indexes. Below, we shall present the methodology of an important comparison of the correlation between the structure of scientific research, scientometric indicators, and GDP of several countries from the EU and outside the EU [98].

The methodology is based on the following indicators and indexes of research production of the scientific community of a country:

1. **Journal paper citedness**

$$\text{JPC} = \frac{C}{P}, \quad (3.56)$$

where

- P: number of journal papers produced by the research community in a country for the time interval of interest;
- C: number of citations obtained by the researchers from the scientific community of a country for the time interval of interest.

2. **Relative subfield citedness**

$$\text{RW} = \frac{C}{P[C/P]_{st}}, \quad (3.57)$$

where

- P: number of journal papers produced by the research community in a country for the time interval of interest and for the research field of interest;
- C: number of citations obtained by the scientists from the scientific community of a country for the time interval of interest and for the scientific field of interest.

- $[C/P]_{st}$: Journal paper citedness for the corresponding field in the world (obtained by the data from a large database such as Web of Science or Scopus).

3. **Journal paper productivity**

$$\text{JPP} = \frac{P}{Pop}, \tag{3.58}$$

where

- P: number of journal papers of a country;
- Pop: population of the country in millions of people.

4. **Highly cited papers productivity**

$$\text{HCPP} = \frac{HCP}{Pop}, \tag{3.59}$$

where

- HCP: number of highly cited papers (ranking among the top 1 % most cited for their subject field and year of publication);
- Pop: population of the country in millions of people.

5. **Relative prominence index**

$$\text{RPI} = \left(\frac{P_c}{\sum P_c}\right) / \left(\frac{P}{\sum P}\right), \tag{3.60}$$

where

- $\frac{P_c}{\sum P_c}$: share of cited papers of a country within the total number of papers cited in the world;
- $\frac{P}{\sum P}$: share of journal papers of a country within the total number of papers in the world.

6. Specific impact contribution

$$\text{SIC} = \frac{C\%}{P\%}, \tag{3.61}$$

where

- $C\%$: percentage share of citations of a country within the total number of citations in the world;
- $P\%$: percentage share of a country in journal papers within the total number of papers in the world.

7. **Rate of highly cited researchers**

$$\text{RHCR} = \frac{HCR}{Pop}, \tag{3.62}$$

where

- HCR: number of researchers of a country in the top 1 % of the researchers most cited;
- Pop: population of the country in millions of people.

8. **Composite publication index**

$$\text{CPI} = w_1(JPP) + w_2(SIC) + w_3(HCPP), \tag{3.63}$$

where

- n: number of countries in the world;
- $w_1 = 1/\sum_{i=1}^{n} JPP_i$;
- $w_2 = 2/\sum_{i=1}^{n} SIC_i$;
- $w_3 = 3/\sum_{i=1}^{n} HCPP_i$

9. **Field structure difference index for country k in field i**

$$\text{FSD}_{k,i} = \frac{(P_{k,i} - P_{s,i})^2}{P_{s,i}}, \tag{3.64}$$

where

- $P_{k,i}$: is the percentage share of publications of country k in the ith scientific field.
- $P_{s,i}$: is the mean percentage share of the standard. As the standard one considers fourteen European Community member states (member states that are not from Eastern Europe) plus the USA and Japan.

10. **Mean structural difference index**

$$\text{MSD}_k = \frac{1}{F}\sum_{i=1}^{F} \frac{(P_{k,i} - P_{s,i})^2}{P_{s,i}}, \tag{3.65}$$

where i is the number of considered scientific subfields.

Vinkler [98] applied the above procedure to the EU countries and to several other countries. We note here that the differences among the countries are well exhibited by the values of the mean structural difference index. For example, the value of this index for Germany for 1995–2005 was 0.18; for the USA, 0.68; for the Czech Republic, 1.17; and for Bulgaria, 2.25. And while the Czech Republic has moved close to the fourteen West European countries, the structure of science in Bulgaria differs greatly from the standard (provided by the fourteen EU countries plus the USA and Japan).

3.13 Efficiency of Research Production from the Point of View of Publications and Patents

As countries become more developed, the ratio between paper production and patent production changes [99]. And the ratio between produced papers and produced patents normalized by population of the country can be considered an index of efficiency of the corresponding national research system. This methodology is developed further in [100]. An analysis of a country's efficiency (within some group of countries) can be made the basis of the following indexes:

1. **Patents–papers index**

$$E_1 = \frac{Pat}{Pap}, \tag{3.66}$$

where

- Pat: number of patents per one million inhabitants of the country;
- Pap: number of papers per one million inhabitants of the country.

2. **Expenditure efficiency index**

$$E_2 = \frac{GERD}{Pap}, \tag{3.67}$$

where

- Pap: number of papers written by the country's researchers;
- $GERD$: gross expenditure on research and development.

3. **Manpower efficiency index**

$$E_3 = \frac{Pap}{MP}, \tag{3.68}$$

where

- Pap: number of papers written by the country's researchers;
- MP: manpower (number of people participating in research activities).

4. **Patent expenditure efficiency index**

$$E_4 = \frac{GERD}{Pat}, \tag{3.69}$$

where

- Pat: number of patents obtained by the country's researchers;
- $GERD$: gross expenditure on research and development.

5. **Patent manpower efficiency index**

$$E_5 = \frac{Pat}{MP}, \tag{3.70}$$

where

- Pat: number of patents obtained by the country's researchers;
- MP: manpower (number of people participating in research activities).

An analysis of several countries performed in [100] shows low efficiency in publishing but high efficiency in patenting in the USA. This pattern is observed also for Germany, Japan, France, and Korea, and China is moving to join this club.

3.14 Indicators for Leadership

Indicators for leadership can be used to assess institutional and national publication activities. Klavans and Boyack [101] consider three kinds of indicators for leadership.

1 *Indicators for current leadership*: Current leadership indicators are connected to the count of the current research publications. These indicators refer to research groups, research institutions, or countries that lead in terms of numbers of papers published, particularly if attention is paid to the most current literature [102].
2. *Indicators for discovery leadership*: These indicators refer to research groups, research institutions, or countries that lead in terms of any of a number of impact measures, which are typically based on citation counts to older literature. For example, a nation with a larger fraction of highly cited papers in a particular field may be considered a discovery leader in the corresponding research field [103]. Other indicators for discovery leadership may be the total citations and fraction of the top one percent of highly cited papers for the earlier time period. One has to be careful, since citation levels can be artificially inflated due to self citations. Special attention should be given to negative citations, which may indicate problems in the corresponding research.
3. *Indicators for thought leadership*: These indicators are a bridge between current leadership and discovery leadership. Thought leadership is an activity measure that examines whether current papers are building on more recent discoveries or on older discoveries in a field. An indicator for thought leadership is the mean reference date in the list of references of the published articles. Thought leadership shows the research groups, institutions, or countries that are quick to follow recent discoveries, e.g., a research group with mean reference date 2012 is quicker to follow research discoveries in comparison to a research group whose mean reference date is 1999. A research group, research organization, or country is considered a thought leader if it is building on the more recent discoveries in its field. At the national policy level, the measure of thought leadership should be age of the scientific environments that the nation wants to pursue [101]. At

this level, the nations that are thought leaders fund mostly young research areas. But even in young research areas, there are discoveries that are of different ages. This is connected to thought leadership at the group (laboratory) level. At this level, where the choice of topic is given, the measure shifts to relative age. Thus when an area of science is targeted, the scientists from groups that are thought leaders focus on the most recent discoveries within this area. Then a country may be a thought leader in some research (i.e., the most recent research areas for this kind of research are funded), but the research groups in this country may not be thought leaders in the corresponding research (if they focus on discoveries that are not the latest in the corresponding research areas).

3.15 Additional Characteristics of Scientific Production of a Nation

Schubert and Braun [104] considered the following relative indexes of scientific production of researchers from different nations and scientific fields (the indexes can be applied also to scientific organizations within a country):

1. **Activity index**
   This index was proposed in [105] and further studied in [106]. It is defined as follows:
   $$AI = \frac{N_1}{N_2}, \tag{3.71}$$
   where
   - N_1: the given field's share in the country's publication output;
   - N_2: the given field's share in the world's publication output.

   $AI = 1$ means that the country's research effort in a given scientific field corresponds to the world average; $AI > 1$ means that the country's effort is greater than the world's average effort.
   Instead of the world average, one can use the average with respect to a set of countries of interest. In this case, the activity index becomes
   $$AI^* = \frac{N_1}{N_2^*}, \tag{3.72}$$
   where
   - N_1: the given field's share in the country's publication output;
   - N_2^*: the given field's share in the publication output of the selected set of countries.

   On the basis of the activity index, one can introduce the *relative specialization index*

$$RSI = \frac{AI - 1}{AI + 1}. \tag{3.73}$$

The relative specialization index has values from -1 to 1 inclusive. $RSI = -1$ means that there is no activity in the corresponding research field. $RSI = 1$ arises when no field other than the given one is active. Negative values of RSI indicate activity that is lower than the average activity. Positive values of RSI indicate activity that is higher than average activity. $RSI = 0$ means that the country's research effort in a given scientific field corresponds to the world average.
The relative specialization index gives evidence of the existence of four patterns in the national publication profiles of the countries of the world [5]:

- *The Western model*: the characteristic pattern of the developed Western countries with clinical medicine and biomedical research as dominating fields;
- *The Japanese model*: engineering and chemistry are dominant. This model is typical also for other developed Asian economies;
- *The former socialist countries model*: physics and chemistry are dominant. Such a model may be observed in the East-European countries, Russia, and China;
- *The bio-environmental model*: biology and earth and space sciences are dominant. Such a model is observed in Australia, South Africa, and some developing countries with relatively large territory and natural resources.

2. **Attractivity index**

$$AAI = \frac{N_3}{N_4}, \tag{3.74}$$

where

- N_3: the given field's share in the citations attracted by the country's publications;
- N_4: the given field's share in the citations attracted by all publications in the world.

This index can be reformulated to compare a country to a set of other countries:

$$AAI^* = \frac{N_3}{N_4^*}, \tag{3.75}$$

where

- N_3: the given field's share in the citations attracted by the country's publications;
- N_4^*: the given field's share in the citations attracted by all publications in the selected set of countries.

3. **Relative citation rate**
   This index is defined as

$$RCR = \frac{N_5}{N_6}, \tag{3.76}$$

   where

   - N_5: observed citation rate over all papers published by the given country in the given field;
   - N_6: observed citation rate over all papers published by the selected set of countries in the given field.

   Observed citation rate of a paper is the actual citation rate and *expected citation rate of a paper* is the average citation rate of the journal in which the paper has been published.

   $RCR > 1$ means that the papers produced by the scientists of a country in the scientific field of interest are more frequently cited than the standard citation rate, and $RCR < 1$ means that the papers are less frequently cited than expected (one reason for this (among many reasons) may be related to their quality).

On the basis of the activity and attractivity indexes, one can produce a *relational chart* of countries (or of scientific organizations in a country). The relational chart is produced as follows: The value of the activity index appears on the x-axis; and the value of the attractivity index appears on the y-axis. The diagonal is the line where the observed and expected citation rates match exactly. If a point corresponding to a country is below the diagonal (and far from the diagonal), this is a sign of problems. A significant distance of a point from the diagonal means that AI or AAI differ significantly for 0. There is a test to check whether the difference is significant [104]:

1. One calculates

$$t_{AI} = \frac{AI - 1}{\Delta_{AI}}; \quad t_{AAI} = \frac{AAI - 1}{\Delta_{AAI}}, \tag{3.77}$$

   where

$$\Delta_{AI} = AI\sqrt{1/N - 1/S}; \quad \Delta_{AAI} = AAI\sqrt{1/M - 1/T},$$

   and

   - N: number of country's publications in the given field;
   - M: number of country's citations in the given field;
   - S: number of country's publications in all scientific fields;
   - T: number of country's citations in all scientific fields;

2. if $t < 2$, the corresponding indicator does not differ significantly from 1 at a significance level of 0.95.

An analogous test can also be performed for the relative citation rate. First one calculates

$$t_{RCR} = \frac{RCR - 1}{\Delta_{RCR}}, \tag{3.78}$$

where

$$\Delta_{RCR} = \sqrt{RCR\frac{Q}{N}}$$

and

- N: country's publications in the given field;
- Q: solution of the equation $\frac{\ln Q}{Q-1} = -\frac{\ln f}{X}$, where X is the mean observed citation rate per publication and f is the fraction of uncited publications.

Then if $t_{RCR} < 2$, RCR does not differ significantly from 1 at a significance level of 0.95.

On the basis of the RCR index, one can introduce another index that rewards papers with RCR value larger than 1 and "punishes" papers with RCR smaller than 1 [107]. This index is just

$$RCR_2 = (RCR)^2. \tag{3.79}$$

We shall finish our discussion of production of researchers from a nation with a description of a set of indexes for measurement of scientific production [108] called FSS-indexes ("Fractional Scientific Strength" indexes). These indexes are based on a measurement of average yearly labor production of researchers at various levels of units (individual, field, discipline, entire organization, region, country). The FSS-indexes connect the salary of researchers with results of their research measured by publications and citations.

The FSS-indexes at different levels are

1. **Individual level**

$$FSS_R = \frac{1}{S_R}\frac{1}{t}\sum_{i=1}^{N} f_i \frac{c_i}{\overline{c}}, \tag{3.80}$$

where

- S_R: average yearly salary of researcher;
- t: number of years of work of researcher in the period of observation;
- N: number of publications of researcher in the period of observation;
- f_i: fractional contribution of researcher to publication i;
- c_i: citations received by the ith publication;
- $\overline{c}$: average number of publications received for all cited publications of the same year and subject category.

2. **Research field level**

$$FSS_F = \frac{1}{S_F} \sum_{i=1}^{N} f_i \frac{c_i}{\overline{c}}, \tag{3.81}$$

where

- S_F: total salary of the research staff (working in the corresponding research field) in the observed period;
- N: Number of publications of the above research staff in the period of observation;
- f_i: fractional contribution of researchers form evaluated group to publication i;
- c_i: citations received by the publication i;
- $\overline{c}$: average number of publications received for all cited publications of the same year and subject category.

3. **Department level**

$$FSS_D = \frac{1}{N_{RS}} \sum_{i=1}^{N_{RS}} \frac{FSS_{R_i}}{\overline{FSS_R}}, \tag{3.82}$$

where

- N_{RS}: number of researches in the department for the observed period;
- FSS_{R_i}: productivity of the ith researcher from the department for the observed period;
- $\overline{FSS_R}$: average national productivity of all productive researchers from the same scientific discipline.

4. Level of multifield units: Such units, for example, are universities or a system of research institutes or even the entire national research system. In this case,

$$FSS_U = \sum_{i=1}^{N_U} \frac{S_{SD_k}}{S_U} \frac{FSS_{SD_k}}{\overline{FSS_{SD_k}}}, \tag{3.83}$$

where

- S_U: total salary of the research staff of the multifield unit for the observed period;
- S_{SD_k}: total salary of the research staff from the observed unit that works in the scientific discipline k in the observed period of time.
- N_U: number of scientific disciplines in the observed unit;
- FSS_{SD_k}: labor productivity in the scientific discipline SD_k of the evaluated unit;
- $\overline{FSS_{SD_k}}$: weighted average of the research productivities in all other units of the kind of unit that is evaluated (of all other universities if the evaluated unit is a university)

The FSS-indexes could lead to quite interesting results for research units and countries where the salaries of researchers are low and their scientific production is not very low. Then it can happen that the effectiveness of the research units in such countries is very good.

3.16 Brief Remarks on Journal Citation Measures

Journal citation measures are much used in library science, research evaluation, etc. In research evaluation, the journal citation measures are applied at all levels: from evaluation of research of individual researchers to evaluation of national research performance. Because of this, we shall mention below several of these measures.

The first very successful journal citation measure was the *impact factor* [109]. The relationship for this index for a journal is

$$IF_n = \frac{c_n}{p_{n-2} + p_{n-1}}, \tag{3.84}$$

where

- c_n: number of citations obtained in the year n by the papers published in the journal in the years $n-1$ and $n-2$;
- p_{n-1}: number of papers published in the journal in the year $n-1$;
- p_{n-2}: number of papers published in the journal in the year $n-2$.

The impact factor is much used today, and it has various strengths such as stability, reproducibility, comprehensibility (the impact factor measures the frequency with which an average article published in a given journal has been cited in a particular year) and independence of the size of the journal (on the number of articles published in the journal per year). In order to be useful, the impact factor must be used carefully, e.g., the impact factors of journals must be used with great care for the purposes of comparison of production of researchers from different scientific areas. One should keep in mind, e.g., that a single measure might not be sufficient to describe citation patterns of scientific journals [5].

In analogy to the impact factor, one may also define the *intermediacy index*

$$II_n = \frac{c_n}{p_n}, \tag{3.85}$$

where

- c_n: number of citations obtained in year n by the papers published in the journal in year n;
- p_n: number of papers published in the journal in year n.

Another index is the *SNIP indicator* (source normalized impact per paper) [110]. The classic version of SNIP is defined as follows:

$$\text{SNIP} = \frac{\text{RIP}}{\text{RDCP}}, \tag{3.86}$$

where

- RIP (raw impact per paper): the RIP value of a journal is equal to the average number of times the journal's publications in the three preceding years were cited in the year of analysis. For example, if 200 publications appeared in a journal in the period 2012–2014 and if these publications were cited 600 times in 2015, then the RIP value of the journal for 2015 equals $600/200 = 3$. What is specific is that *in the calculation of RIP values, citing and cited publications are included only if they have the Scopus document type article, conference paper, or review*. The RIP indicator is similar to the journal impact factor, but the RIP indicator uses three instead of two years of cited publications and includes only citations to publications of selected document types.
- RDCP (relative database citation potential): RDCP is calculated as follows:

$$\text{RDCP} = \frac{\text{DCP}}{\text{m(DCP)}}, \tag{3.87}$$

  where

  – DCP (database citation potential): DCP is calculated as follows:

$$\text{DCP} = \frac{\sum_{i=1}^{n} r_i}{n}, \tag{3.88}$$

    where n is the number of publications in the subject field of the journal and r_i denotes the number of references in the ith publication to publications that appeared in the three preceding years in journals covered by the database.
  – m(DCP): the median DCP value of all journals in the database.

Finally, let us mention the *SJR*: Scimago journal rank, which is based on the transfer of prestige from a journal to another journal [111]. Prestige is transferred through the references that a journal makes to the rest of the journals and to itself. The SJR is calculated as follows:

$$\text{SJR}_j = \frac{1-d-e}{N} + \frac{e\,\text{Art}_i}{\sum_{j=1}^{N}\text{Art}_i} + d\sum_{j=1}^{N}\frac{C_{ji}\text{SJR}_j}{C_j}\frac{1-\left[\sum_{k\in\{\text{Dangling nodes}\}}\right]}{\sum_{h=1}^{N}\sum_{k=1}^{N}\frac{C_{kh}\text{SJR}_k}{C_k}} + d\left[\sum_{k\in\{\text{Dangling nodes}\}}\right]\frac{\text{Art}_i}{\sum_{j=1}^{N}\text{Art}_j}, \tag{3.89}$$

where

- C_{ij}: citations from journal j to journal i.
- C_j: number of references of journal j.
- N: number of journals.
- d: constant (usually equal to 0.85).
- e: constant (usually equal to 0.1).
- Art_i: number of articles in journal i.
- Dangling nodes: these are journals of the universe that do not have references to any other journal of the universe, although they can be cited or not. They constitute impasses in a graph, since from them it is impossible to jump to other nodes. In order to ensure that the iterative process is convergent, dangling nodes are virtually connected to all those of the universe, and its prestige is distributed between all the nodes proportionally to the number of articles of each.

On the basis of the SDJR, one can calculate another index specific to the ith journal:

$$\text{SJRQ}_i = \frac{\text{SJR}_i}{\text{Art}_i}. \tag{3.90}$$

The iterative procedure of calculation of the SJR involves the following three steps:

1. Initial assignment of the SJR: a default prestige is assigned to each journal. The calculation of the SJR is a converging process, so the initial values don't determine the final result (but the initial values influence the number of iterations needed).
2. Iteration process of calculation: departing from step 1, the computation is iterated to calculate the prestige of each journal based on the prestige transferred by the rest. The process ends when the variation of the SJR between two iterations is less than a limit fixed before the calculation process. The final result is the SJR of each journal.
3. Computation of SJRQ: After the computation of SJR of all journals, one divides the SJR by the number of articles published in the citation window. The result is the average prestige per article.

Another version of the SJR (the SJR2) is also available [112]. Let us note that a major drawback of the journal impact factor is its lack of field (subject) normalization, i.e., differences in citation volumes between different fields are not taken into account. SNIP belongs to indexes that are based on the idea that citations to publications should be normalized with respect to the length of the reference lists of the citing publications (sources). The source normalized indexes are based on the observation that the reference lists' lengths vary across fields. Source-normalized indexes do not require a field classification scheme. There are also indexes based on other ideas. An example is MNCS (mean normalized citation score) [113, 114], based on the approach to field normalization, in which a classification scheme is used (i.e., each publication is assigned to one or more of the fields of the scheme). In the case of

MNCS, citation scores of the target publications (e.g., the publications under evaluation) are compared to expected citation scores for publications in the fields to which the publications belong (these fields are the Thomson Reuters subject categories of journals).

3.17 Scientific Elites. Geometric Tool for Detection of Elites

Elites are very important parts of social structures [115–117]. There exist characteristic features of research organizations that lead to the formation of research elites. Usually a small number of researchers publish many papers and a small number of researchers are highly cited. These categories of researchers form some of the scientific elites. Elites are of great importance for the dynamics and evolution of scientific structures and systems. Because of this, scientific elites are the subject of intensive research [118–129].

There is a **square root law of Price** [130]: *half of the literature on a subject will be contributed by the square root of the total number of authors publishing in that area.*

Let $g(x)$ represent the probability of an author making x published contributions to a subject field. Then the mathematical formulation of the square root law of Price is [131]

$$\lim_{x_{\max}\to\infty}\left[\frac{\sum_{x=h}^{x_{\max}} x\ g(x)}{\sum_{x=1}^{x_{\max}} x\ g(x)}\right]=\frac{1}{2}, \tag{3.91}$$

where h is such that

$$\left[\sum_{x=1}^{x_{\max}} g(x)\right]^{1/2}=\sum_{x=h}^{x_{\max}} g(x). \tag{3.92}$$

Let the total number of authors in a scientific discipline be A. The law of Price can be generalized as follows [78]: A^{α} *authors will generate a fraction* α *of the total number of papers*. Then if $\alpha = 1/2$, one obtains the square root law of Price.

One can select groups of elite researchers on the basis of the law of Price. Another kind of possible rule for selecting an elite is the arithmetic $a\%/b\%$-rule: $a\%$ *of the papers are produced by* $b\%$ *of the scientists*. The most famous of these rules is the 80/20-rule: 80 % of the papers are produced by 20 % of the scientists. (Note that it is not necessary that $a + b = 100$.)

In the next chapter we shall discuss more of the theory of Price for scientific elites. This theory will lead us to the following conclusion: assuming the validity of the law of Lotka for scientific publications, one can obtain that the scientific elite consists of scientists whose number of publications is between $0.749\sqrt{i_{max}}$ and i_{max} publications (where i_{max} is the maximum number of publications written by a scientist from the corresponding group of scientists). And the size of this elite is about $\frac{0.812}{\sqrt{i_{max}}}$ of the size of the group of scientists. In this chapter we shall discuss another methodology for

determination of classes of scientific elites. This methodology is based on geometry and doesn't require validity of some law for scientific production. The corresponding measures will be obtained on the basis of the Lorenz curve for the ownership of scientific publications. As we have mentioned above, the Lorenz curve is an instrument for visualization of inequality in a population. It is very popular in the study of wealth distribution in a population [132–134]. Below, we shall be interested in the number of publications owned by researchers from some population (in our case, the population will consist of the members of a research institute). *We note that the measures of the sizes of the elites discussed below can be applied not only to populations of researchers but also to all populations that can be characterized by a Lorenz curve. Thus the methodology discussed below may be used to determine elites with respect to other characteristics of scientific production, such as the number of citations.*

3.17.1 Size of Elite, Superelite, Hyperelite, …

Let us consider the Lorenz curve shown in Fig. 3.1. Let us trace the diagonal from the point (0, 1) to the point (1, 0) in the (P, L)-plane. This diagonal crosses the Lorenz curve at a point with coordinates $(P_e, 1 - P_e)$. We shall consider the number $1 - P_e$

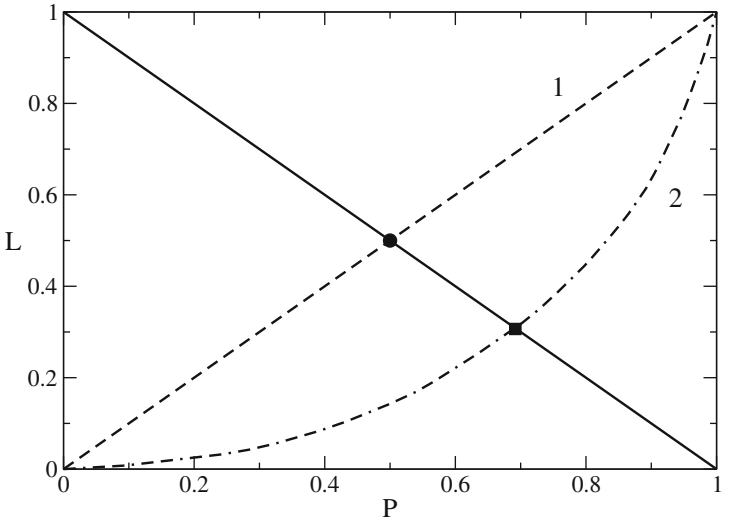

Fig. 3.1 Elite size measure by the Lorenz curve. The measure is the coordinate $1 - P_e$ of the cross point of the diagonal $(P, 1 - P)$ and the corresponding Lorenz curve. For the Lorenz curve marked by 1 (all scientists own the same number of papers), the cross point (*filled circle*) has coordinates (0.5, 0.5). In percentages, this is the 50/50-curve (nonelite distribution). For the Lorenz curve marked by 2 (corresponding to the situation at the Institute of Mechanics of the Bulgarian Academy of Sciences), the cross point (*filled square*) is (0.69, 0.31). In percentages, this is the 69/31 curve

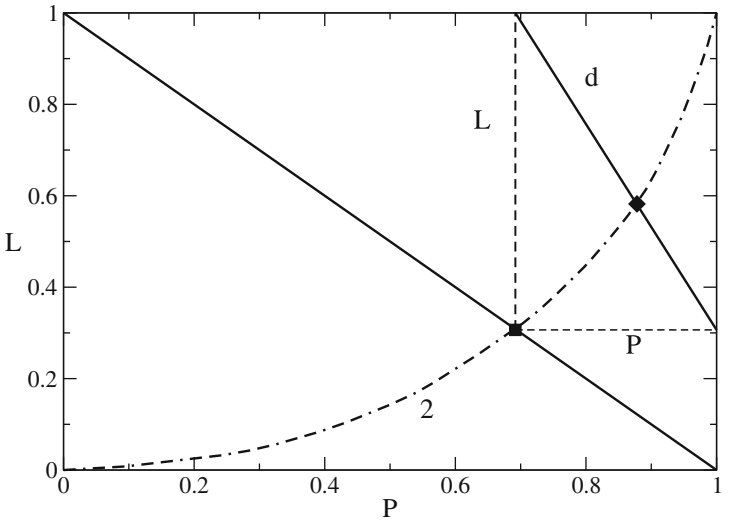

Fig. 3.2 The geometric measure for the scientific superelite by the Lorenz curve. The Lorenz curve marked by 2 is the same as in Fig. 3.1. One introduces a new Cartesian coordinate system with axes P^* and L^* and initial point that coincides with the point $(P_e, 1 - P_e)$ connected with the definition of the size of the scientific elite from Fig. 3.1. In this new coordinate system, the diagonal marked with d is plotted. The point (P_s, L_s) marked by a diamond gives the size and the production of the corresponding superelite. For the case of the Lorenz curve 2 (corresponding to the Institute of Mechanics of the Bulgarian Academy of Sciences), the coordinates of the point marked by a diamond are approximately $(P_s, L_s) = (0.88, 0.58)$, which means that the corresponding superelite consists of $1 - P_s = 0.12$, i.e., 12 % of the population of scientists owns $1 - L_s = 0.42$, i.e., 42 % of all papers. We recall here that the measure of the size of the elite from the previous figure tells us that the size of the elite of the institute was 31 % of the scientists, and this elite owns 69 % of the papers produced by the institute scientists

to be a measure of the size of the elite of the population corresponding to the Lorenz curve. Let us discuss this measure a bit further.

For the Lorenz curve corresponding to the case that all scientists own the same number of publications (in this case, the Lorenz curve is the diagonal that connects the points (0, 0) and (1, 1)), we have $P_e = 0.5$. We shall call such a curve a curve of class 50/50 (the elite has its maximum size). We can continue the construction of geometric measures one step further, and this will lead us to the concept of the scientific superelite. The procedure is illustrated in Fig. 3.2. The next step (definition of the superelite and its size) is geometrically analogous to the step that led us to the geometric measure of the size and production of the scientific elite. For this step, the initial point of the Cartesian coordinate system is not $(P, L) = (0, 0)$ but $(P, L) = (P_e, 1 - P_e)$, where P_e is the coordinate connected to the point corresponding to the geometric elite measure above (i.e., the point that is the intersection point of the Lorenz curve and the diagonal marked with a solid line in Fig. 3.2). Next we construct the axes P^* and L^* shown in Fig. 3.2. Finally, we plot the diagonal d shown

in Fig. 3.2, and the intersection point of this diagonal with the Lorenz curve gives us the geometric measure of the size and the production of the superelite. This point is marked with a diamond in Fig. 3.2, and its coordinates can be easily calculated. The coordinates of the point marked by a square (let us call it point E), which gives the size and production of the elite, are $E = (P_e, 1 - P_e)$. Then the coordinates of the point marked by a diamond (let us call it point S) are $S = (P_s, P_e \frac{P_s - P_e}{1 - P_e})$. For the case of the 61/39 curve marked by 2 in Fig. 3.2 and $P_s = 0.88$ measured by the intersection of the Lorenz curve and diagonal d, we obtain the coordinates of the point S to be approximately $S = (0.88, 0.58)$. In summary:

1. **Elite**: the coordinate P_e gives us information about the size and production of the scientific elite of the group of scientists described by the corresponding Lorenz curve.
2. **Superelite**: The coordinates P_e and P_s give us information about the size and production of the corresponding superelite.
3. **Hyperelite**: We can continue the process of construction of geometric measures starting now from the point S. What we shall obtain is the next point (let us call it H), which shall give us information about a smaller group of scientists called the hyperelite. The coordinates of this point will be $(P_h, \frac{P_h - P_s}{1 - P_s})$. Then the coordinates P_e, P_s, P_h will give us information about the size and production of the hyperelite.

The above geometric procedure may be continued further, and additional higher-order elites may be determined.

3.17.2 Strength of Elite

Next we can introduce a quantity that we shall call strength of the elite. Let us consider a geometric measure connected to the size and production of the elite. This measure is connected to the point E that has coordinates $(P_e, L_e = 1 - P_e)$. We define the strength of the elite as

$$s_e = \frac{1 - L_e}{1 - P_e} = \frac{P_e}{1 - P_e}. \tag{3.93}$$

We can define also the strength of the superelite. The coordinates of the point S connected to the size and production of the superelite are $S = (P_s, L_s)$. Then the strength of the superelite is defined as

$$s_s = \frac{1 - L_s}{1 - P_s} = \frac{1 - P_e \frac{P_s - P_e}{1 - P_e}}{1 - P_s} = \frac{1 - P_e(1 + P_s - P_e)}{(1 - P_e)(1 - P_s)}. \tag{3.94}$$

Finally, we can define the relative size of the superelite with respect to the size of the corresponding elite:

$$S_{se} = \frac{1 - P_s}{1 - P_e}. \tag{3.95}$$

Table 3.1 Parameters of the scientific elites and superelites of the studied institutes of the Bulgarian Academy of Sciences. $1 - P_e$: size of the scientific elite; $1 - L_e$: percentage of total number of papers owned by the members of the scientific elite; s_e: strength of the scientific elite; $1 - P_s$: size of the scientific superelite; $1 - L_s$: percentage of total number of papers owned by the members of the scientific superelite; s_e: strength of the scientific superelite. The studied institutes are from Bulgarian Academy of Sciences: Institute of Mathematics and Informatics (IMI); Institute of Mechanics (IMECH); Institute of Information and Communication Technologies (IIKT); Institute of Solid State Physics (ISSP); Institute of Electronics (IE); Institute of Optical Materials and Technologies (IOMT); Institute of Nuclear Research and Nuclear Energy (INRNE); Central Laboratory for Solar Energy and New Energy Sources (CLSENES)

Institute	$1 - P_e$ (%)	$1 - L_e$ (%)	s_e	$1 - P_s$ (%)	$1 - L_s$ (%)	s_s
IMI	34	64	1.88	14	38	2.71
IICT	30	70	2.33	12	40	3.33
IMECH	31	69	2.23	12	42	3.50
CLSENES	32	68	2.13	14	39	2.79
IOMT	35	65	1.86	16	35	2.19
IE	32	68	2.13	13	41	3.15
ISSP	34	66	1.88	14	39	2.79
INRNE	32	68	2.13	13	40	3.08

We note that the measures discussed above are different from the classic measures connected to the scientific elites.

Table 3.1 shows results about the size and production of the elites and superelites at the studied institutes of the Bulgarian Academy of Sciences. The sizes and productivities are very close, which means that in the size–production plane, the elites and the superelites form two clusters of researchers.

The elites at the mathematics and the physics institutes consist of about one-third *of the scientists, and these elites own about* two-thirds *of the scientific publications of the corresponding institute. The superelites consists of about* one-seventh *of the scientists, and they own about* two-fifths *of the scientific production. In addition, about two-thirds of the scientists do not belong to the scientific elites, and all these scientists own about one-third of the scientific production of the corresponding institute. Six-sevenths of the scientists do not belong to the superelite, and these scientists own about three-fifths of the scientific production of the corresponding institute.*

After selection of researchers that belong to elite, superelite, etc., one can study different characteristics of the selected groups of researchers. Here we shall mention only one of these characteristics: the age structure of the studied Bulgarian elites and superelites. Almost 80 % of the members of the superelites are of age 60 and older. In ten years, these scientists will no longer be staff scientists of the corresponding institute. Such people are also in the majority of the corresponding elites. The younger generation of scientists (ages between 40 and 60) is insufficiently represented in the scientific elites and scientific surepelites. Hence entire fields of national scientific research can be under the influence of aging researchers. This may have negative consequences, since many cases, the growth rate of research production is positive

and increases up to the ages about 30. After that age, the growth rate of research productivity usually begins to decrease [135]. This effect may not concern scientists belonging to superelites and hyperelites. And when such a researcher is no longer active, this is a great loss to the national research program in the corresponding research field.

References

1. M. Gibbons, C. Limoges, H. Nowotny, S. Schwartzman, P. Scott, M. Trow, *The new production of knowledge: the dynamics of science and research in contemporary societies* (Sage, London, 1994)
2. L.K. Hessels, H. van Lente, Re-thinking new knowledge production: a literature review and a research agenda. Res. Policy **37**, 740–760 (2008)
3. R.A. Boschma, Proximity and innovation: a critical assessment. Reg. Stud. **39**, 61–74 (2005)
4. I. Rafols, Knowledge integration and diffusion: measures and mapping of diversity and coherence, ed. by Y. Ding, R. Rousseau, D. Wolfram, *Measuring Scholarly Impact. Methods and Practice*. (Springer, Cham, 2014), pp. 169–192
5. W. Glänzel, *Bibliometrics as a research field: a course on theory and application of bibliometric indicators* (Ungarische Akademie der Wissenschaften, Budapest, 2003)
6. P. Brown, The half-life of the chemical literature. J. Am. Soc. Inform. Sci. **31**, 61–63 (1980)
7. R.E. Burton, R.W. Kebler, The "half-life" of some scientific and technical literatures. Am. Documentation **11**, 18–22 (1960)
8. P. Vinkler, *The Evaluation of Research by Scientometric Indicators* (Chandos, Oxford, 2010)
9. P. Vinkler, Publication velocity, publication growth and impact factor: an empirical model, ed by B. Cronin, H.B. Atkins. *The Web of Knowledge: A Festschrift in Honor of Eugene Garfield*. ASIS Monograph Series (Information Today Inc, Medford, NJ, 2000), pp. 163–176
10. P. Vinkler, Research contribution, authorship and team cooperativeness. Scientometrics **26**, 213–230 (1993)
11. A. Przeworski, (Institutionalization of voting patterns or is mobilization the source of decay? Am. Polit. Sci. Rev. **69**, 49–67 (1975)
12. R.R. Shutz, On the measurement of income inequality. Am. Econ. Rev. **41**, 107–122 (1951)
13. A.B. Atkinson, On the measurement of inequality. J. Econ. Theory **2**, 244–263 (1970)
14. F.A. Cowell, *Measuring Inequality* (Oxford University Press, Oxford, UK, 2011)
15. P.D. Allison, Measures of inequality. Am. Sociol. Rev. **43**, 865–880 (1978)
16. A.R. Wilcox, Indices of qualitative variation and political measurement. Western Political Quart. **26**(2), 325–343 (1973)
17. A.L. Wilcox, *Indices of Qualitative Variation*. ORRN-TM-1919, (Oak Ridge National Laboratoty, Oak Ridge, Tennessee, 1967)
18. S.S. Nagel, *Public Policy: Goals, Means and Methods* (St. Martin Press, New York, 1984)
19. A.P. Lüthi, *Messung wirtschaflicher Ungleichheit.* Lecture Notes in Economic and Mathematical Systems No. 189 (Springer, Berlin, 1981)
20. C. Gini, *Variabilita e mutabilita* (Bologna, Italy, 1912)
21. L. Ceriani, P. Verme, The origins of Gini index: extracts from variabilita e Mutabilita (1912) by Corrado Gini. J. Econ. Inequality **10**, 421–443 (2012)
22. H.G.P. Jansen, Gini's coefficient of mean difference as a measure of adoption speed: theoretical issues and empirical evidence from India. Agric. Econ. **7**, 351–369 (1992)
23. I.I. Eliazar, I.M. Sokolov, Measuring statistical evenness: a panoramic overview. Phys. A **391**, 1323–1353 (2012)
24. S. Yitzhaki, E. Schechtman, *The Gini Methodology* (Springer, New York, 2013)

25. J.G. Rodriguez, R. Salas, The Gini coefficient: majority voting and social welfare. J. Econ. Theory **152**, 214–223 (2014)
26. B. Milanovic, A simple way to calculate the Gini coefficient, and some implications. Econ. Lett. **56**, 45–49 (1997)
27. C.J. Groves-Kirkby, A.R. Denman, P.S. Phillips, Lorenz curve and Gini coefficient: novel tools for analysing seasonal variation of environmental radon gas. J. Environ. Manage. **90**, 2480–2487 (2009)
28. J. Yang, X. Huang, X. Liu, An analysis of education inequality in China. Int. J. Educ. Dev. **37**, 2–10 (2014)
29. K. Kimura, A micro-macro linkage in the measurement of inequality: another look at the Gini coefficient. Qual. Quant. **28**, 83–97 (1994)
30. P.A. Rogerson, The Gini coefficient of inequality: a new interpretation. Lett. Spatial Resour. Sci. **6**, 109–120 (2013)
31. M.-H. Huang, H.-H. Chang, D.-Z. Chen, The trend in scientific research and technological innovation: a reduction of the predominant role of the U.S. in world research and technology. J. Infometrics **6**, 457–468 (2012)
32. A. Stirling, A general framework for analysing diversity in science, technology and society. J. Royal Soc. Interface **4**, 707–719 (2007)
33. A.O. Hirschman, *National Power and Structure of Foreign Trade* (University of California Press, Berkeley, CA, 1945)
34. O.C. Herfindahl, *Concentration in the steel industry*. Ph.D. Thesis, (Columbia University, 1950)
35. R. Linda, Competition policies and measures of dominant power, ed. by H.W. de Jorg. W.G. Shepherd, *Mainstreams in Industrial Organization* (Martinus Nijhoff Publishers, Dordrecht, 1986), pp. 287–307
36. G. Chammas, J. Spronk. Concentration measures in portfolio management, ed. by S. Greco, B. Bouchoin-Meunter, G. Colleti, M. Fedrizzi, B. Matarazzo, R.R. Yager, Advances in Computational intelligence, in *14th International Conference on Information Processing and Management of Uncertainty in Knowledge-based Systems*, IPMU 2012, Catania, Italy, July 9–13, 2012, Proceedings, Part IV. (Springer, Berlin, 2012), pp. 94–103
37. A. Arlandis, E. Baranes, Interactions between network operators, content producers and internet intermediaries: empirical implications on network neutrality. Intereconomics **2**, 98–105 (2011)
38. W. Naude, R. Rossouw, Export diversification and economic performance: evidence from Brazil, China, India and South Africa. Econ. Change Restructuring **44**, 99–134 (2011)
39. J. Horvath, Suggestion for a comprehensive measure of concentration. South. Econ. J. **36**, 446–452 (1970)
40. J.L. Ray, D. Singer, Measuring the concentration of power in the international system. Sociol. Methods Res. **1**, 403–437 (1973)
41. R. Taagepera, J.L. Ray, A generalized index of concentration. Sociol. Methods Res. **5**, 367–383 (1977)
42. S. Lieberson, Measuring population diversity. Am. Sociol. Rev. **34**, 850–862 (1969)
43. L.A. Renzulli, H. Aldrich, Who can you turn to? The activation within core business discussion networks. Soc. Forces **84**, 323–341 (2005)
44. J.R. Bond, The influence of constituency diversity on electoral competition in voting for Congress 1974–1978. Legislative Stud. Quart. **8**, 201–217 (1983)
45. C.R. Rao, Diversity and dissimilarity coefficients: a unified approach. Theor. Popul. Biol. **21**, 24–43 (1982)
46. C. Ricotta, L. Szeidl, Towards a unifying approach to diversity measures: bridging the gap between the Shannon entropy and Rao's quadratic index. Theor. Popul. Biol. **70**, 237–243 (2006)
47. L. Cassi, W. Mescheba, E. Turckheim, How to evaluate the degree of interdisciplinarity of an institution? Scientometrics **101**, 1871–1895 (2014)

48. O.D. Duncan, B. Duncan, A methodological analysis of segregation indexes. Am. Sociol. Rev. **20**, 210–217 (1955)
49. R. Taagepera, Inequality, concentration, imbalance. Polit. Methodol. **6**, 275–291 (1979)
50. D.W. Rae, M. Taylor, *The Analysis of Political Cleavages* (Yale University Press, New Haven, Conn, 1971)
51. L. Leydesdorff, Indicators of structural change in the dynamics of science: entropy statistics of the SCI Journal Citation Reports. Scientometrics **53**, 131–159 (2002)
52. H. Theil, The desired political entropy. Am. Polit. Sci. Rev. **63**, 521–525 (1969)
53. H. Theil, On the estimation of relationships involving qualitative variables. Am. J. Sociol. **76**, 103–154 (1970)
54. Y. Wang, Decomposing the entropy index of racial diversity: in search of two types of variance. Ann. Reg. Sci. **48**, 897–915 (2012)
55. K.D. Bailey, Sociological entropy theory: toward a statistical and verbal congruence. Qual. Quant. **18**, 113–133 (1983)
56. B. Raj, J. Koerts (eds.), *Henri Theil's Contributions to Economics and Econometrics*. Volume 2: Consumer demand analysis and information theory. (Kluwer, Dordrecht, 1992)
57. H. Theil, *Economics and Information Theory* (North Holland, Amsterdam, 1967)
58. D.F. Batten, *Spatial Analysis of Intercatting Economies* (Kluwer, Dordrecht, 1983)
59. K. Kesselman, French local politics: a statistical examination of grass roots consensus. Am. Polit. Sci. Rev. **60**, 963–974 (1966)
60. H. Theil, *Statistical Decomposition Analysis* (North Holland, Amsterdam, 1972)
61. J. Fellman, Lorenz curve. ed by M. Lovric. *International Encyclopedia of Statistical Science* (Springer, Berlin, 2011), pp. 760–761
62. D. Chotikapanich, *Modeling Income Distributions and Lorenz Curves* (Springer, New York, 2008)
63. E. Scalas, T. Radivoevic, U. Garibaldi, Wealth distribution and Lorenz curve: a finitary approach. J, Econ. Interact. Coordinartion (in press) (2015). doi:10.1007/s11403-014-0136-2
64. G. Warner, A Lorenz curve based index of income stratification. Rev. Black Polit. Econ. **28**, 41–57 (2001)
65. J. Tang, X. Wang, Analysis of land use structure based on Lorenz curves. Environ. Monit. Coord. **151**, 175–180 (2009)
66. O. Alonso-Villar, Measuring concentration: Lorenz curves and their decompositions. Ann. Reg. Sci. **47**, 451–475 (2011)
67. P. Suppes, Lorenz curves for various processes: a pluralistic approach to equity. Soc. Choice Welfare **5**, 89–101 (1988)
68. L. Egghe, Conjugate partitions in infometrics: Lorenz curves, h-type indices, Ferrer graphs and Durfee squares in a discrete and continuous setting. J. Infom. **4**, 320–330 (2010)
69. R. Rouseau, Measuring concentration: sampling design issues, as illustrated by the case of perfectly stratified samples. Scientometrics **28**, 3–14 (1993)
70. L. Egghe, R. Rousseau, Symmetric and asymmetric theory of relative concentration and applications. Scientometrics **52**, 261–290 (2001)
71. L. Egghe, R. Rousseau, How to measure own-group preference? A novel approach to a sociometric problem. Scientometrics **59**, 233–252 (2004)
72. R. Ketzer, K.F. Zimmermann, Publications: German scientific institutions on track. Scientometrics **80**, 231–252 (2009)
73. S. Shibayama, Distribution of academic research grants: a case of Japanese national research grant. Scientometrics **88**, 43–60 (2011)
74. B. Jarneving, Regional research and foreign collaboration. Scientometrics **83**, 295–320 (2010)
75. W. Halffman, L. Leydesdorf, Is inequality among universities increasing? Gini coefficients and the elusive rise of elite universities. Minerva **48**, 55–72 (2010)
76. T.J. Cleophas, A.H. Zwinderman, Pareto charts for identifying the main factors of multifactorial outcomes. ed. by T.J. Cleophas, A.H. Zwinderman. *Machine Learning in Medicine* (Springer, Berlin, 2014), pp. 101–106

77. S.H. Kan, *Metrics and Models in Software Quality Engineering* (Addison-Wesley, Boston, 2002)
78. L. Egghe, R.A. Rousseau, A characterization of distributions which satisfy Price's law and consequences for the laws of Zipf and Mandelbrot. J. Inform. Sci. **12**, 193–197 (1986)
79. S. Lieberson, Rank-sum comparisons between groups, ed. by D. Heise. *Sociological Methodology* (Jossey-Bass, San Francisco, 1976), pp. 276–291
80. S. Lieberson, An asymmetrical approach to segregation, ed. by C. Peach, V. Robinson, S. Smith. *Ethnic Segregation in Cities* (Croom Helm, London, 1981), pp. 61–82
81. N. Toren, V. Kraus, The effects of minority size on women's position in academia. Soc. Forces **65**, 1090–1100 (1987)
82. M. Fosset, J.S. Scott, The measurement of intergroup income inequality: a conceptual review. Social Forces **61**, 855–871 (1983)
83. P.B. Coulter, Measuring the inequity of urban public services. Policy Stud. J. **8**, 683–698 (1980)
84. M.T. Marsh, D.A. Schilling, Equity measurement in facility location analysis: a review and framework. Eur. J. Oper. Res. **74**, 1–17 (1994)
85. K. Barker, The UK research assessment exercise: the evolution of a national research evaluation system. Res. Eval. **16**, 3–12 (2007)
86. G. Falavigna, A. Manello, External funding, efficiency and productivity growth in public research: the case of the Italian National Research Council. Res. Eval. **23**, 33–47 (2014)
87. B.M. Coursey, A.N. Link, Evaluating technology-based public institutions: the case of radiopharmaceutical standards research at the National Institute of Standards and Technology. Res. Eval. **7**, 147–157 (1998)
88. G. Lewison, Evaluation of national biomedical research outputs through journal-based esteem measures. Res. Eval. **5**, 225–235 (1995)
89. C.M. Sa, A. Kretz, K. Sigurdson, Accountability, performance assessment, and evaluation: policy pressures and responses from research councils. Res. Eval. **22**, 105–117 (2013)
90. F. Xu, X.X. Li, W. Meng, W.B. Liu, J. Mingers, Ranking academic impact of world national research institutes014by the Chinese Academy of Sciences. Res. Eval. **22**, 337–350 (2013)
91. L. Georghiou, Research evaluation in European national science and technology systems. Res. Eval. **5**, 3–10 (1995)
92. N. Kastrinos, Y. Katsoulacos, Towards a national system of research evaluation in Greece. Res. Eval. **5**, 63–68 (1995)
93. C.-G. Yi, K.-B. Kang, Developments of the evaluation system of government-supported research institutes in Korean science and technology. Res. Eval. **9**, 158–170 (2000)
94. M. Coccia, A basic model for evaluation R&D performance: theory and application in Italy. R&D Manage. **31**, 453–464 (2001)
95. M. Coccia, Models for measuring the research performance and identifying the productivity of public research institutes. R&D Manage. **34**, 267–280 (2005)
96. M. Coccia, A scientometric model for the assessment of scientific research performance within public institutes. Scientometrics **65**, 307–321 (2005)
97. M. Coccia, Measuring performance of public research units for strategic change. J. Infometrics **2**, 184–194 (2008)
98. P. Vinkler, Correlation between the structure of scientific research, scientometric indicators and GDP in EU and non-EU countries. Scientometrics **74**, 237–254 (2008)
99. E. Albuquerque, Science and technology systems in less developed countries, ed. by H. Moed, W. Glaänzel, U. Schmoch. *Handbook of Quantitative Science and Technology Research* (Kluwer , Dordrecht, 2005), pp. 759–778
100. A. Basu, The Albuquerque model and efficiency indicators in national scientific productivity with respect to manpower and funding of science. Scientometrics **100**, 531–539 (2014)
101. R. Klavans, K. Boyack, Thought leadership: a new indicator for national and institutional comparison. Scientometrics **75**, 239–250 (2008)
102. A.F.J. van Raan, Statistical properties of bibliometric indicators: research group indicator distributions and correlations. J. Am. Soc. Inform. Sci. Technol. **57**, 408–430 (2006)

103. D.A. King, The scientific impact of nations. Nature **430**, 311–316 (2004)
104. A. Schubert, T. Braun, Relative indicators and relational charts for comparative assessment of publication output and citation impact. Scientometrics **9**, 281–291 (1986)
105. J.D. Frame, Mainstream Research in Latin America and the Caribbean. Interciencia **2**, 143–147 (1977)
106. R. Rousseau, L. Yang, Reflections on the activity index and related indicators. J. Infometrics **6**, 413–421 (2012)
107. P. Vinkler, Weighted impact on publications and relative contribution score. Two new indicators characterizing publication activity of countries. Scientometrics **14**, 161–163 (1988)
108. G. Abramo, C.A. D'Angelo, How do you define and measure research productivity? Scientometrics **101**, 1129–1144 (2014)
109. E. Garfield, *Citation Indexing: Its Theory and Applications in Science, Technology and Humanities* (Wiley, New York, 1979)
110. H.F. Moed, Measuring contextual citation impact of scientific journals. J. Informetrics **4**, 265–277 (2010)
111. B. Gonzalez-Pereira, V.P. Guerrero-Bote, F. Moya-Anegon, A new approach to the metric of journals scientific prestige: the SJR indicator. J. Informetrics **4**, 379–391 (2010)
112. V.P. Guerrero-Bote, F. Moya-Anegon, A further step forward in measuring journals' scientific prestige: the SJR2 indicator. J. Informetrics **6**, 674–688 (2012)
113. L. Waltman, N.J. van Eck, T.N. van Leeuwen, M.S. Visser, A.J.F. van Raan, Towards a new crown indicator: an empirical analysis. Scientometrics **87**, 467–481 (2011)
114. L. Waltman, N.J. van Eck, T.N. van Leeuwen, M.S. Visser, A.J.F. van Raan, Towards a new crown indicator: some theoretical considerations. J. Informetrics **5**, 37–47 (2011)
115. R. Abzug, Community elites and power structure, ed. by R.A. Cnaan, C. Milofsky, *Handbook oF Community Movements and Local Organizations* (Springer, New York, 2008), pp. 89–101
116. J.S. Coleman, *Power and Structure of Society* (Norton, New York, 1974)
117. L. Trilling, Technological elites in France and the United States. Minerva **17**, 225–243 (1979)
118. N. Elias, H. Martins, R. Whitley, *Scientific Establishments and Hierarchies* (Reidel, Dordrecht, 1982)
119. M. Mulkay, The mediating role of the scientific elite. Soc. Stud. Sci. **6**, 445–470 (1975)
120. H. Best, U. Becker, *Elites in Transition*. Elite research in Central and Eastern Europe. (VS Verlag für Sozialwissenschaften, 1997)
121. H. Zuckerman, *Scientific Elites*. Nobel laureates in the United States. (Free Press, New York, 1977)
122. J.N. Parker, C. Lortie, S. Allesina, Characterizing a scientific elite: the social characteristics of the most highly cited scientists in environmental science and ecology. Scientometrics **85**, 129–143 (2010)
123. M. Davis, C. Wilson, Elite researchers in ophthalmology: aspects of publishing strategies, collaboration and multi-disciplinarity. Scientometrics **52**, 395–410 (2001)
124. E. Lazega, L. Mounier, M.-T. Jourda, R. Stofer, Organizational vs. personal social capital in scientists' performance: a multi-level network study of elite French cancer researchers (1996–1998). Scientometrics **67**, 27–44 (2006)
125. C. Cao, R.P. Suttmeier, China's new scientific elite: distinguished young scientists, the research environment and hopes for Chinese science. China Quart. **168**, 960–984 (2001)
126. R.S. Hunter, A.J. Oswald, B.G. Charlton, The elite brain drain. Econ. J. **119**, F231–F251 (2009)
127. G. Laudel, Migration currents among scientific elite. Minerva **43**, 377–395 (2005)
128. B. Golub, The Croatian scientific elite and its socio-professional roots. Scientometrics **43**, 207–229 (1998)
129. N.C. Mullins, Invisible colleges as scientific elites. Scientometrics **7**, 357–368 (1985)
130. D. De Solla Price, *Little Science, Big Science* (Columbia University Press, New York, 1963)
131. W. Glänzel, A. Schubert, Price distribution: an exact formulation of Price's 'square root law'. Scientometrics **7**, 211–219 (1985)

132. J.L. Gast, The estimation of the Lorenz curve and Gini index. Rev. Econ. Stat. **54**, 306–316 (1972)
133. N.C. Kakwani, Applications of Lorenz curves in economic analysis. Econometrica: J. Econometric Soc. **43**, 719–727 (1977)
134. A. Dragulescu, V.M. Yakovenko, Exponential and power-law probability distributions of wealth and income in the United Kingdom and the United States. Phys. A **299**, 213–221 (2001)
135. A. van Heeringen, P.A. Dijkwel, The relationships between age, mobility and scientific productivity. Part II. Effect of age on productivity. Scientometrics **11**, 281–293 (1987)