

Citation expectations: are they realized? Study of the Matthew index for Russian papers published abroad

Vladimir Pislyakov

Higher School of Economics, Library, 101000, Myasnitskaya, 20, Moscow
(Russia)
pislyakov@hse.ru

Ekaterina Dyachenko

Higher School of Economics, Perm Branch, Library, 614070,
Studencheskaya, 38, Perm (Russia)
dyachenkoel@hse.perm.ru

Abstract We consider the “Matthew effect” in the citation process which leads to reallocation (or misallocation) of the citations received by scientific papers within the same journals. The case when such reallocation correlates with a country where an author works is investigated. Russian papers in chemistry and physics published abroad were examined. We found that in both disciplines in about 60% of journals Russian papers are cited less than average ones. However, if we consider each discipline as a whole, citedness of a Russian paper in physics will be on the average level, while chemistry publications receive about 16% citations less than one may expect from the citedness of the journals where they appear. Moreover, Russian chemistry papers mostly become undercited in the leading journals of the field. Characteristics of a “Matthew index” indicator and its significance for scientometric studies are also discussed.

Keywords Matthew index; Citedness; International comparison; Physics; Chemistry

Introduction

The term “Matthew effect” was introduced by Robert K. Merton (Merton, 1968). He described a psychosocial mechanism that led to misallocation of credit in the reward system of science. Papers written by eminent scholars (e. g. Nobel laureates) tend to get disproportionately great credit while relatively unknown scientists tend to get disproportionately little credit for contributions of the same quality. The name for the effect comes from the Gospel saying “for unto every one that hath shall be given, and he shall have abundance: but from him that hath not shall be taken away even that which he hath” (Matthew 25:29). Later Merton (1988) developed his ideas further.

A group of German scholars (Bonitz et al. 1997; Bonitz et al. 1999; Bonitz and Scharnhorst 2001) discovered and investigated a similar effect for citation of works written by scientists from different countries. They found that there are several countries that obtain more citations than it could be assumed from the citedness of the journals where they publish their work. At the same time there are a plenty of other countries whose publications are cited less frequently than average papers in the same journals. German researchers also discovered that this redistribution correlates with the impact factors of the journals where a country publishes its works. If a country

has high “citation expectations” (based on the citedness of the journals where it publishes), then it probably will get even more citations than expected. If a country has low citation expectations, its works generally will be even more undercited. This effect was called by Bonitz et al. (1997) “Matthew effect for countries”.

To measure the degree of citation redistribution an indicator named “Matthew index” (MI) was defined as the ratio of the difference between observed (OC) and expected (EC) numbers of citations to the expected number of citations. (We use variable names OC and EC instead of more usual OCR and ECR as the latter imply rates, that is *cites per paper*.) First of all, MI_j for a single journal (j) may be written as

$$MI_j = \frac{OC_j - EC_j}{EC_j}$$

Here expected number of citations (EC_j) is calculated on the basis of the average number of cites per paper in the journal. Namely, if citedness of an average paper in the journal is ECR_j , then $EC_j = p_j \cdot ECR_j$, where p_j is number of papers of a country in journal j . A total number of citations actually received by these papers is OC_j . If Matthew index for some country’s papers in some journal j is positive, it means that these papers are cited more than average paper in this journal, if MI_j is less than zero, then they are undercited compared to journal’s average paper.

Matthew index may then be calculated for some set of journals, e. g. for those pertaining to some scientific field or even for all scientific periodicals. The definition is as follows:

$$MI = \frac{\sum_j OC_j - \sum_j EC_j}{\sum_j EC_j}$$

Where the sums are taken over all journals in a chosen set. Note that MI thus defined differs from the simple sum of the MI_j . Again, if $MI > 0$ then the country’s papers taken as a whole are overcited, if $MI < 0$ they are undercited in the journal set under consideration.

It should be mentioned that “Matthew terminology” goes side by side with that of relative citation indicators as described, for example, by Schubert et al. (1983) or Schubert and Braun (1986). More precisely, relative citation rate (RCR) used there equals $MI + 1$. Conceptually similar indicator was recently introduced by the name “J Factor” (Ball et al. 2009).

We aim at investigating the Matthew index for Russian papers published abroad, i. e. in non-Russian journals. This will reveal how publications of Russian scholars look against papers of their foreign colleagues in the context of the same journals. Bonitz (2002; 2005) emphasized that Matthew index measures “competitiveness” of a nation’s scientific literature. It is undoubtedly true, especially if we keep in mind that Matthew effect for countries is in fact a *redistribution* of citations, and if one country gains positive Matthew index then some others should have negative ones: it is *their* citations that are redistributed in favour of the winner(s).

For our research we chose two scientific fields which are considered traditionally strong for Russian science, chemistry and physics. According to Essential Science Indicators (ESI) ranking these fields have the highest number of papers and receive the greatest number of citations compared to all other Russian sectors of science. Our investigation should show how competitive Russian papers are in chemistry and physics when they are published in prestigious international journals.

Finally, not much attention has recently been paid to such an interesting scientometric aspect of a Matthew effect as “Matthew effect for countries”, though its other manifestations were actively investigated, e. g. (Medoff 2006; Morgan et al. 2008; Larivière and Gingras in print). Our contribution seeks to improve this situation.

Methodology

A Thomson Reuters database Science Citation Index Expanded, a part of Web of Science (WoS) product hosted on the Web of Knowledge online platform, was used to get publication data. Only documents of the type “Article” were taken into account. (Note that data were gathered in spring 2008, while in autumn 2008 some of the “Articles” transformed their document type in Web of Science to “Proceedings Paper”, so that they both must now be included to reproduce our results, see (Thomson Reuters 2009) for details.) A paper was attributed to Russia if country of at least one of the institutional affiliations of its authors was “Russia”. It is the so-called “whole counting” method; note that Bonitz with coauthors used first-author method. Citations are counted in WoS through all citation indexes (including Science Citation Index Expanded, Social Sciences Citation Index, Arts & Humanities Citation Index, Conference Proceedings Citation Index).

Data were collected and Matthew indices were calculated for two disciplines, chemistry and physics. For subject classification of the journals ESI field classification was used. This system classifies all WoS journals into 22 broad fields, “Chemistry” and “Physics” are among them. Only one field is assigned to each journal, so two journal sets do not overlap.

We chose two years of publication, 2003 and 1997. A 4-year citation period was chosen, including the year of publication: citations from 2003–2006 literature were counted for 2003 papers and from 1997–2000 literature for 1997 papers.

To investigate citedness of Russian papers published abroad, we excluded all Russian journals found in the ESI lists. As it is hard to determine “nationality” of the journal in this time of globalization of science, a formal criterion was used: all journals that published in the corresponding year more than 50% articles from Russia were excluded from further analysis. Journals that had not published a single Russian paper during that year were also excluded, as their citedness did not influence Matthew indices for Russia. Four final journal sets for the analysis are called “chemistry-2003”, “chemistry-1997”, “physics-2003” and “physics-1997”. They contain 222, 212, 183 and 164 sources respectively. We will also speak of “chemistry-2003 papers” (papers published in 2003 in chemistry-2003 journals) and so on.

Results and discussion

General characteristics

Quantitative characteristics of the journal/paper sets are summarized in Table 1. More foreign journals in chemistry than in physics publish at least one Russian paper (still, this is partly explained by overall number of WoS chemical journals), however the total number of papers with Russian (co)authors in physics is more than twice higher than that in chemistry.

Table 1. Sources and papers included into analysis

Set	Journals	Papers, total	Russian papers	Share of Russian papers, %
chemistry-2003	222	76359	1718	2.3
chemistry-1997	212	67488	1582	2.3
physics-2003	183	71498	4002	5.6
physics-1997	164	58854	3883	6.6

Figure 1 illustrates the distribution of Russian papers among the journals of these sets. The maximum, minimum and median percentages of the papers with Russian authors are shown. Figure 1 confirms that relative presence of Russian papers is higher in physics and there is slightly negative trend from 1997 to 2003 in both disciplines.

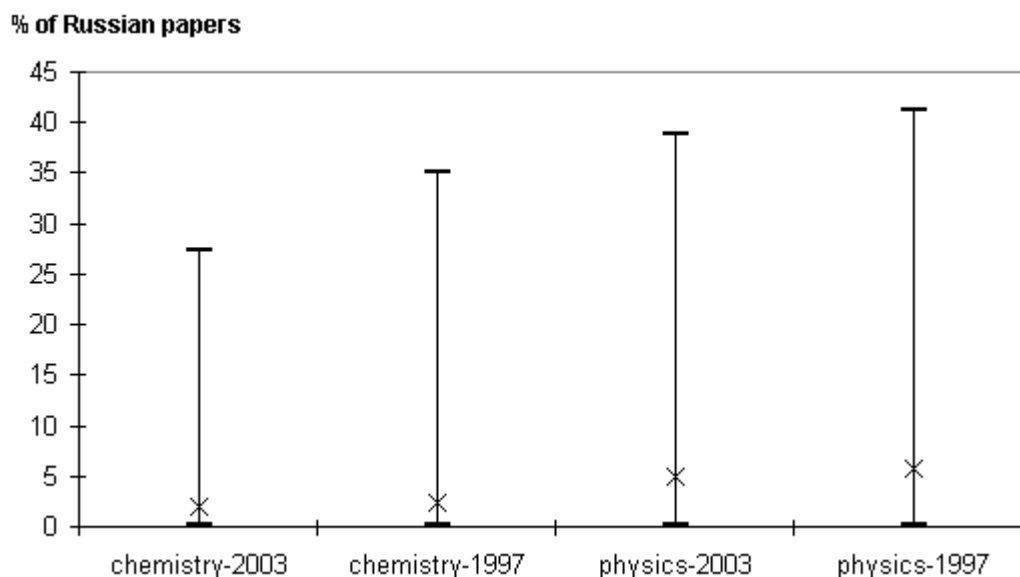


Figure 1. Distribution of Russian papers among journals in the 4 journal sets. Higher and lower dashes represent journals with maximum and minimum percentage of Russian papers in them, a cross shows the median

Citation indicators

Figure 2 shows “gross” citation indicators (without discerning the journal level, no Matthew-index considerations yet) for all 4 sets of papers. A sharp difference between relative standing of chemistry and physics papers published by Russian authors in foreign journals is evident. While an average Russian paper in physics is cited on the world average level, a paper in chemistry obtains only about 70% of citations made to the world’s average paper. (It should be stressed once more that we consider *only non-Russian journals*, so the overall cites per Russian paper would probably be lower if we also included national journals which generally have lower impact factors (Zitt et al. 2003).) These shares are almost constant for the years under consideration, though there is a slight positive movement from 1997 to 2003 in physics and a slight negative shift in chemistry.

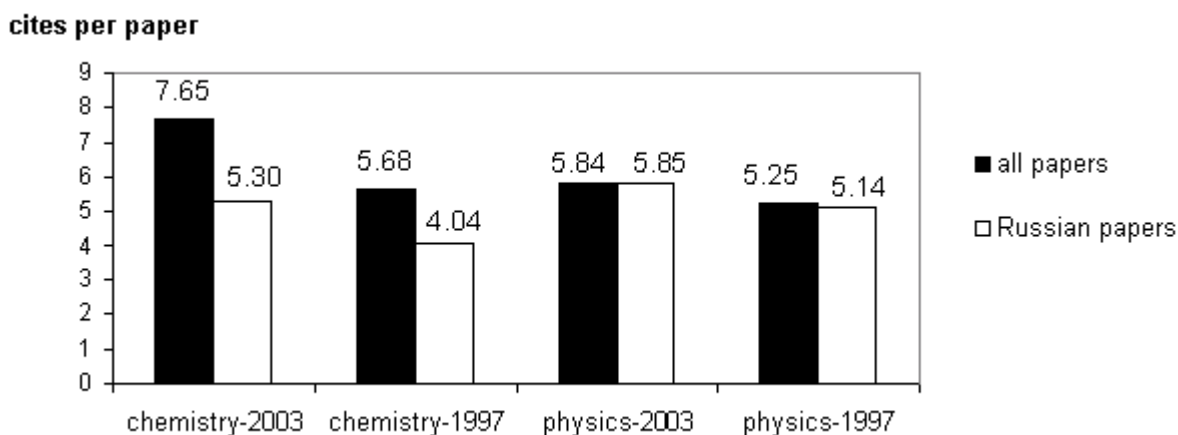


Figure 2. Cites per paper for 4 sets of papers

Another remark that should be made before we proceed to Matthew indicators refers to how the 2003 and 1997 papers are cited in time. This is shown for all sets of papers in Figure 3, with the percentage of citations to each set made in each year. For more broad analysis we added one more year into this graph, so 100% are total 2003-2007 and 1997-2001 numbers of citations to 2003 and 1997 papers, respectively. One may notice that the mode of each distribution occurs at the third year (including the year of publication), so after that the citation curve decreases. The added fifth year confirms that too. So, our four-year analysis covers years with maximum number of citations and we may expect our results to be reliable.

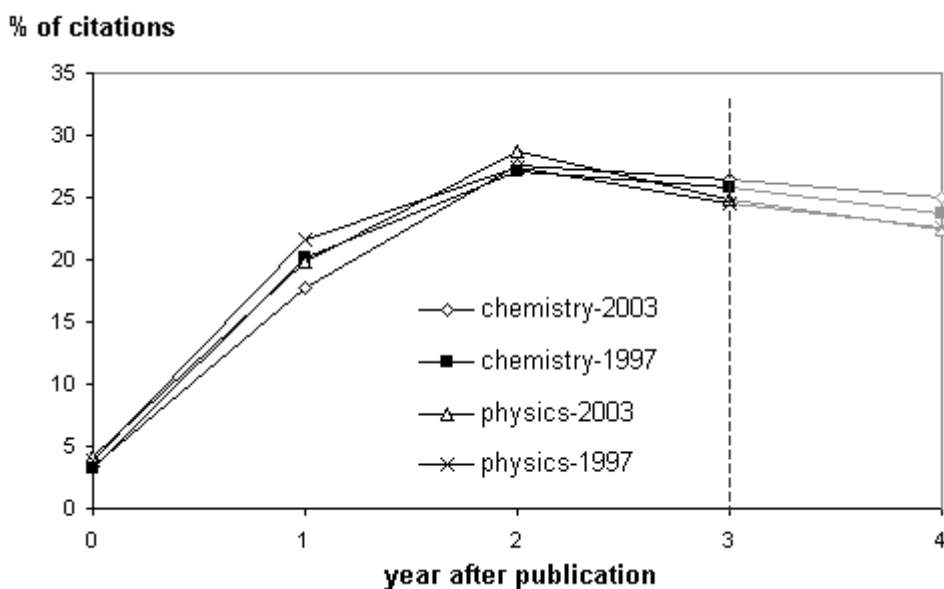


Figure 3. Distribution of citations to paper sets by year. Citations received in 2003-2007 for 2003 sets, in 1997-2001 for 1997 sets. The sum is 100% for the years shown. Dashed line indicates time limit of this study

Matthew indices for journals

Matthew index for Russian papers in a particular journal may be positive (Russian papers are cited more than journal's average) or negative (they are cited less than journal's average).

Distribution of the numbers of journals with positive and negative indices for all 4 sets of sources is shown in Figure 4. For all cases there are more journals where Russian publications are undercited, generally about 60% of the sources have Matthew index less than zero, with the exception of chemistry-2003 set where this share amounts to 68%.

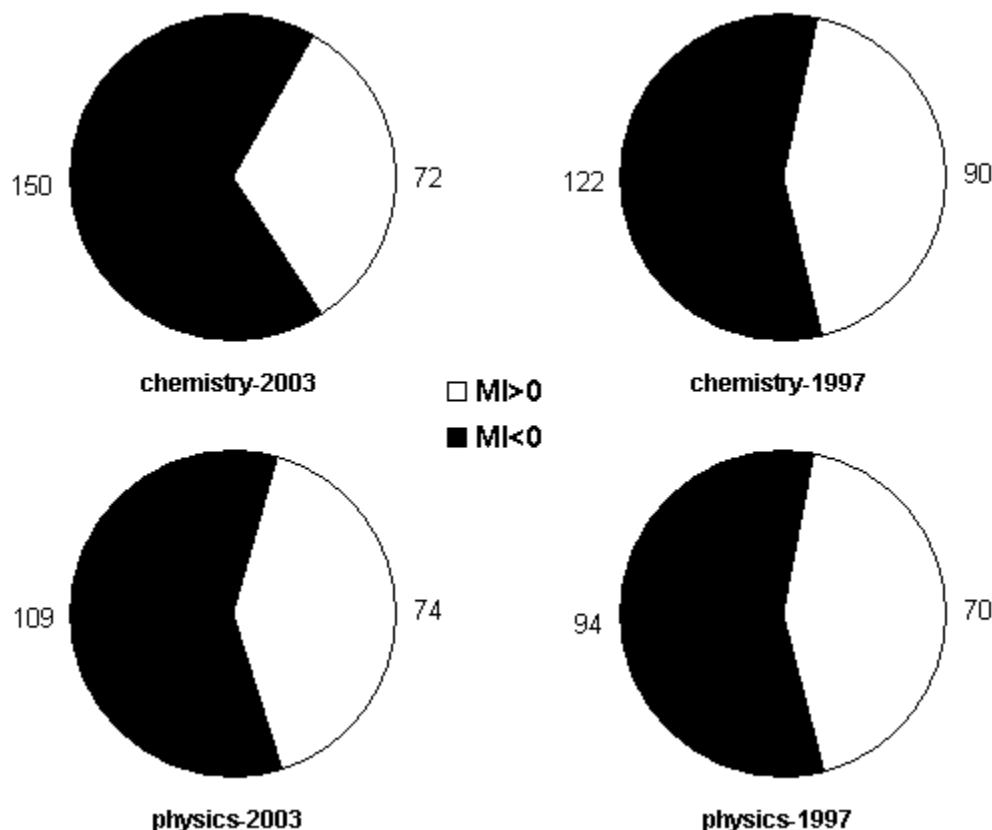


Figure 4. Matthew index signs for Russian papers, by individual journals. Numbers next to the pie charts represent the number of journals with negative and positive values of the Matthew index, respectively

It should be noted that there is nothing extraordinary in the fact that some country's papers are undercited in the majority of journals. Here we bring together a mean indicator (average citedness of a paper in the journal) with a median measure (50% of journals). And, as it is known, in the positively skewed distributions, which citation distributions are supposed to be, the mean is greater than the median. So the majority of countries will probably be undercited in half or more of the journals and it can even be easily constructed an example where *all* countries get their papers undercited in more than 50% of journals.

On the whole, as Table 2 shows, we found no significant correlation between Matthew index for Russian papers in a journal and (a) average citedness of its papers (maximum absolute value of the Pearson correlation coefficient through all 4 sets is 0.21); (b) number of Russian papers in the journal (maximum is 0.10); (c) share of Russian papers in the journal (maximum is 0.08). In a certain sense this is in accordance with Bonitz (2002) who states that number of papers or citations, number of participating countries or impact factor hardly influence the "Matthew core journals" list (journals where papers are more likely to have higher absolute value of Matthew index).

Table 2. Pearson correlation coefficients between Matthew index (*MI*) of individual journals and their other characteristics

Set	<i>MI</i> vs. average citedness of a paper	<i>MI</i> vs. number of Russian papers	<i>MI</i> vs. share of Russian papers
chemistry-2003	-0.13	-0.07	-0.08
chemistry-1997	-0.21	-0.10	-0.05
physics-2003	0.00	-0.10	-0.03
physics-1997	0.02	-0.01	-0.02

Though overall correlation is not significant, it appeared useful to plot “relational charts” (Schubert and Braun 1986) that show how citedness of Russian papers in a journal corresponds to average citedness of a journal’s paper (Figure 5).

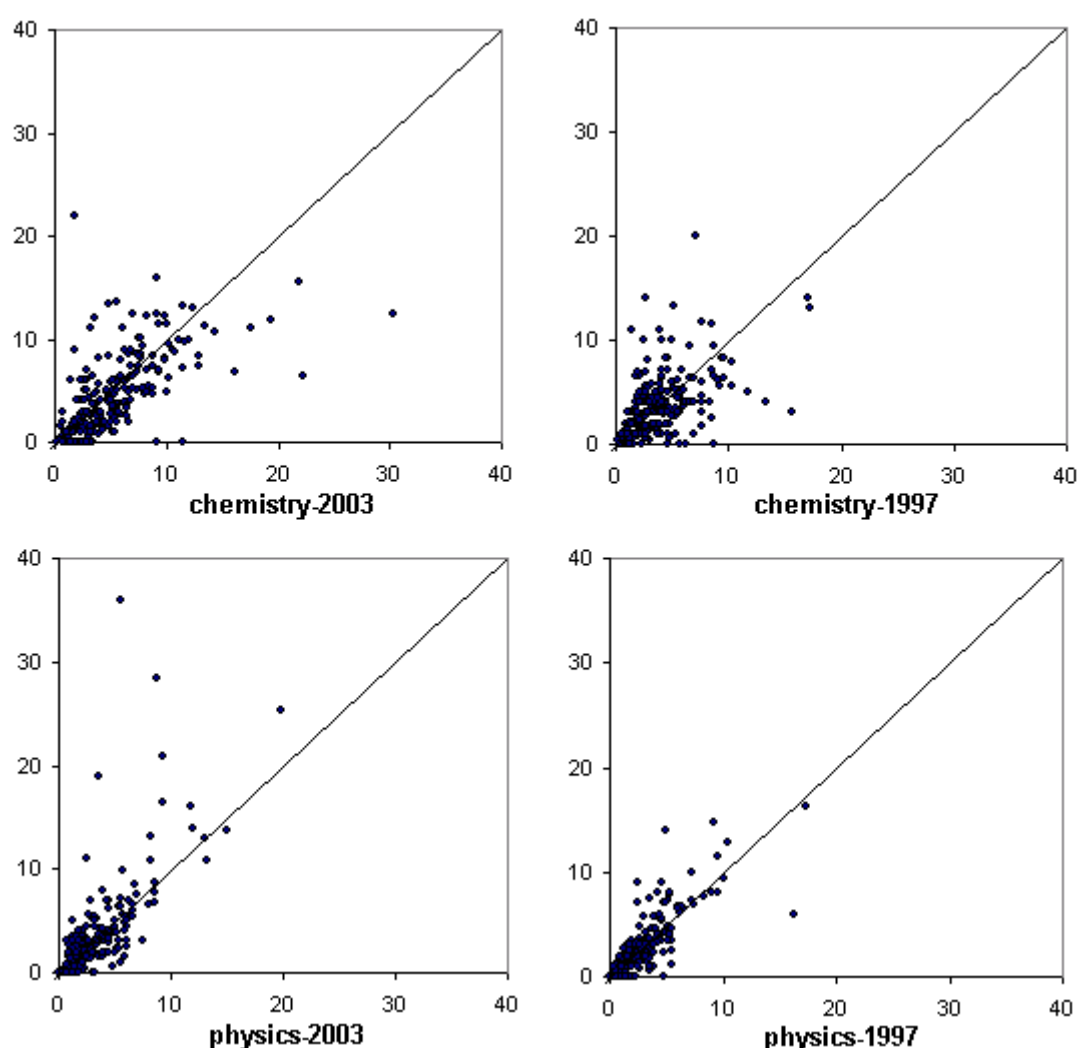


Figure 5. Relational charts for Russian papers.
X-axis — cites per average paper in a journal;
Y-axis — cites per average Russian paper in a journal

The X-axis shows average citedness of a journal’s paper and in fact reflects journal’s “generalized impact factor” in a sense of Rousseau (1988) and Egghe (1988) concept (and

counted for articles only). The Y-axis shows average citedness of Russian papers in the corresponding journal. If Russian publications are cited exactly as an average paper in a journal ($MI = 0$), this journal will lie on the $y=x$ line which is shown on all diagrams. More “observed” citations for Russian papers than “expected” level ($MI > 0$) will move the journal above this line. If Russian papers are undercited ($MI < 0$), this will move the corresponding dots below $y=x$.

There is no clear regularity for dots scattering in Figure 5. Still, some valuable observations may be made. For both chemical sets we may see that in all journals with the highest impact factor (in a generalized sense mentioned above) Russian papers become undercited. The first 10 journals with the most cited papers for chemistry-2003 and 14 journals for chemistry-1997 lie below the $y=x$ line. This means that in the most prestigious chemistry journals Russian papers are regularly cited below average level. The same is not true for physics publications. Among the 10 most cited journals in physics-1997 set 3 have positive Matthew index for Russian papers. For physics-2003 the result is even better: in 8 out of 10 journals with the highest impact (including no. 1, it is *Physical Review Letters*) Russian articles are cited above average level. This distinction in getting credit in the most influential journals marks serious difference between levels of Russian literature in physics and chemistry.

Another interesting question is that of stability of the Matthew index for a given journal through a period of time. On the basis of our statistics for two years of publication for two disciplines no stability was found (Table 3). Journal sets chemistry-1997 and chemistry-2003 share 149 journals in common. Physics-1997 and physics-2003 have 133 overlapping journals. Pearson correlation coefficients for 1997 and 2003 Matthew indices for the same journals are 0.18 for chemistry and 0.11 for physics. It is a very weak correlation, which may suggest that there are no specific journals where Russian papers are systematically undercited and those where Russian papers always receive citations above average level. Additional argument for this is that the index reverses its sign for 69 out of 149 chemistry journals and for 66 out of 133 physics ones. If Russian papers-1997 were undercited in these journals, then Russian papers-2003 are overcited or vice versa.

Table 3. Pearson correlation coefficients between 1997 and 2003 Matthew indices for the same journals

Set	Number of journals common to both sets	Pearson correlation
Chemistry (1997 vs. 2003)		
All journals	149	0.18
Journals with at least 10 Russian papers	28	0.26
Journals with at least 20 Russian papers	14	0.44
Physics (1997 vs. 2003)		
All journals	133	0.11
Journals with at least 10 Russian papers	52	0.06
Journals with at least 20 Russian papers	33	0.31

We also studied this correlation for journals with significant number of Russian papers. The stability of Matthew index through time may be somewhat stronger if we exclude publications with 1 or 2 Russian papers, which appeared there, so to say, “by accident”, as distinguished from journals which regularly publish Russian authors. The effect is observed, the Pearson coefficients generally tend to become greater (Table 3), still even this correlation is not convincing. For example, the coefficients for 1997 and 2003 Matthew indices of 14 chemistry journals with no less than 20 Russian papers equals 0.44, while for 33 such journals in physics it is 0.31.

Matthew indices for disciplines

If we combine the data from all journals in each set, we will get Matthew indices for the whole disciplines, as was described in the Introduction. Table 4 contains these indices as well as the difference between the observed and expected numbers of citations, $OC - EC$ (“surplus”).

Table 4. Matthew index for Russian papers, by disciplines

Set	“Surplus” of citations	Matthew index
chemistry-2003	−1762	−16.2%
chemistry-1997	−723	−10.2%
physics-2003	+160	0.7%
physics-1997	+235	1.2%

We may conclude that Russian chemistry literature receives significantly less citations than it is expected on the basis of citedness of the journals where Russian papers appear. For 2003 articles Russian authors received 1762 citations less than average papers distributed among the same journals. On average, Russian paper in chemistry published in 2003 by the foreign journal gets 16% less citations.

Physics, on the contrary, shows almost exact average level of citedness. Its Matthew index is slightly above zero. The most interesting situation is with physics-1997 set. As we may infer from Figure 2, Russian papers in this set receive as a whole 2% citations below average level. However, Matthew index for them is positive. This demonstrates difference between “gross-” and “micro-level” (journal level) of citation analysis. On the journal level Russian papers were cited better than average papers in the same journal, but, most probably, distribution of papers by Russian authors skewed in 1997 towards sources with lower impact factor, so the overall citedness of Russian publications appeared to be below average level. This effect is closely linked with the concept of “relative publication strategy” (Vinkler 1997).

It may be added that international collaboration strongly influences the Matthew index. Separate calculations for papers written *only* by Russian authors (no other country mentioned in all addresses of the authors’ affiliations) discovered that their Matthew index falls to -32% for chemistry-2003 and to -31% for physics-2003. In many cases when Russian scientists publish their work abroad they need coauthors from other countries to have their papers cited on the average journal level or higher. Perhaps one of the reasons why chemistry papers have lower Matthew index than physics ones is a greater share of the “pure Russian” publications in chemistry (41% against 29% in physics for 2003 sets). Probably the lower overall relative citedness of Russian chemistry papers (Figure 2) also has its explanation, at least in part, in the levels of international collaboration in these fields.

Conclusion

We have studied visible traces of subtle processes in the universe of science communication that lead to reallocation of credit and redistribution of citations to scientific works. This results in non-zero “Matthew index” and both micro-structure and macro-structure of such redistribution were considered. On the micro-level, considering separate journals, the effect was recorded as for Russian papers in chemistry so in physics. In about 60% of foreign journals Russian publications are cited less than an average paper.

As for the macro-view, on the discipline (field) level, Matthew index appeared to be negative for chemistry and slightly above zero for physics. This finding allowed us to suggest that there exists an important difference in the state of these sciences in Russia, or at least in the course of their internationalization. A strong support to this provides the fact that Russian chemistry papers, unlike physics ones, are undercited in all most prestigious journals. Lack of competitiveness of a country's papers in the leading publications in the field should be considered as a significant drawback.

Matthew index proved to be a special, independent indicator that does not correlate with any other bibliometric characteristics of the studied journals. The intriguing question whether the observed undercitedness of some of the Russian literature is a pure psychosocial phenomenon, or it reflects a real difference in value of this corpus of works, deserves more thorough examination in the future. It is perhaps the most complex problem when we are speaking of reallocation of credit, because "natural experiment", such as studied by Larivière and Gingras (in print), may rarely be performed here.

In closing it is worth mentioning that studies related to the Matthew effect in bibliometrics are not a merely theoretical activity, but can bring changes to science policy procedures. For example, Bordons et al. (2002) recommend to use impact factors of journals where a paper is published and not the observed citations for scientometric analysis of "peripheral country's" research. This is to avoid Matthew-like effects and "sociological" component that influences the allocation of scientific reward. Anyhow, two main methods to measure a merit of a scientific paper are counting the citations received by it and considering the impact factor of a journal that has published it. In scientometrics Matthew index stands in an important position, serving as a link between one method and another.

Acknowledgments This research was supported by the Higher School of Economics Scientific Foundation (Grants 08-01-0050 "How papers of Russian scientists published abroad are cited? Study of the Matthew Effect" and 09-01-0046 "Patterns of international coauthorship of Russian scientists: papers and citations"). Authors would like to thank Andrei V. Poletaev, Yuri Khripin and Mikhail Polyanski for their helpful comments on the earlier version of this paper.

References

- Ball, R., Mittermaier, B., Tunger, D. (2009). Creation of journal-based publication profiles of scientific institutions — A methodology for the interdisciplinary comparison of scientific research based on the J-factor, *Scientometrics*, 81, 381-392.
- Bonitz, M. (2002). Ranking of Nations and Heightened Competition in Matthew Core Journals: Two Faces of the Matthew Effect for Countries, *Library Trends*, 50, 440-460.
- Bonitz, M. (2005). Ten years Matthew Effect for Countries, *Scientometrics*, 64, 375-379.
- Bonitz, M., Bruckner, E., & Scharnhorst, A. (1997). Characteristics and Impact of the Matthew Effect for Countries, *Scientometrics*, 40, 407-422.
- Bonitz, M., Bruckner, E., & Scharnhorst, A. (1999). The Matthew Index – Concentration Patterns and Matthew Core Journals, *Scientometrics*, 44, 361-378.
- Bonitz, M. & Scharnhorst, A. (2001). Competition in Science and the Matthew Core Journals, *Scientometrics*, 51, 37-54.
- Bordons, M., Fernandez, M. T., & Gomez, I. (2002). Advantages and limitations in the use of impact factor measures for the assessment of research performance in a peripheral country, *Scientometrics*, 53, 195-206.

- Egghe, L. (1988). Mathematical relations between impact factors and average number of citations, *Information Processing & Management*, 24, 567-576.
- Larivière, V., Gingras, Y. (in print). The impact factor's Matthew effect: a natural experiment in bibliometrics, *Journal of the American Society for Information Science and Technology*. Retrieved September 18, 2009 from <http://arxiv.org/ftp/arxiv/papers/0908/0908.3177.pdf>
- Medoff, M. H. (2006). Evidence of a Harvard and Chicago Matthew effect, *Journal of Economic Methodology*, 13, 485-506.
- Merton, R. K. (1968). The Matthew Effect in Science. *Science*, 3810, 56-63.
- Merton, R. K. (1988). The Matthew Effect in Science, II: Cumulative Advantage and the Symbolism of Intellectual Property, *ISIS*, 79, 606-623.
- Morgan, P. L., Farkas, G., & Hibel, J. (2008). Matthew effects for whom? *Learning Disability Quarterly*, 31, 187-198.
- Rousseau, R. (1988). Citation distribution of pure mathematics journals. In L. Egghe & R. Rousseau (Eds.), *Informetrics 87/88* (pp. 249-262). Amsterdam: Elsevier.
- Schubert, A., & Braun, T. (1986). Relative indicators and relational charts for comparative assessment of publication output and citation impact, *Scientometrics*, 9, 281-291.
- Schubert, A., Glänzel, W., & Braun, T. (1983). Relative citation rate: A new indicator for measuring the impact of publications. In D. Tomov & L. Dimitrova (Eds.), *Proceedings of the First National Conference with International Participation on Scientometrics and Linguistics of Scientific Text* (pp. 80-81). Varna.
- Thomson Reuters (2009). *Why the number of articles in Web of Science has gone down, and the number of proceedings papers has gone up*. Retrieved April 9, 2009 from: http://isiwebofknowledge.com/products_tools/multidisciplinary/webofscience/cpci/usingproceedings/
- Vinkler, P. (1997). Relations of relative scientometric impact indicators. The relative publication strategy index, *Scientometrics*, 40, 163-169.
- Zitt, M., Ramanana-Rahary, S., Bassecoulard, E. (2003). Correcting glasses help fair comparisons in international science landscape: Country indicators as a function of ISI database delineation, *Scientometrics*, 56, 259-282.