Effects of European Union Funding and International Collaboration on Estonian Scientific Impact

Tanel Hirv

Faculty of Economics and Business Administration, University of Tartu, J. Liivi 4, 51009 Tartu, ESTONIA.

ABSTRACT
A positive influence of international collaboration on the impact of research has been previously extensively described. This paper takes a step further by providing an investigation of the effects of funding sources on Estonian research impact based on Thomson Reuters’ citation indexes. We ask whether also European Union (EU) funding in addition to international collaboration help Estonia achieve a higher scientific impact. The present paper uses funding acknowledgement (FA) section included in Web of Science (WoS) for determining sources of funding. For this purpose, articles with Estonia in the address section are selected and retrieved from 2008 to 2015 and are divided into four categories based on their funding sources; national; EU; national and EU simultaneously; and other. Results show that EU funding increases Estonian scientific impact significantly. Although there is some variability between research areas, EU funding combined with international collaboration produces the most cited scientific articles. It suggests that EU funding can help Estonia get a better outcome in international collaboration than otherwise possible. The main limitations of this paper include methodical problems how funding agencies are determined in WoS and the time dependence of citations what makes an evaluation of recent publications robust.

Key words: Web of Science, Estonia, Scientific impact, Funding sources, Research collaboration.

INTRODUCTION
The ability to estimate a nation’s scientific impact is vital for managers who have to make decisions about funding and set research priorities. European countries recognise that for further developments a targeted research policy with thorough studies of efficiency is necessary.[1] This also applies to the European Union (EU) and Estonia. In the context of this paper, we measure scientific impact in terms of citations. Eugenie Garfield, the pioneer of scientometrics states[2] that the total number of citations is about the most objective measure there is of the material’s importance to current research. In addition, citations:[3] 1) constitute a measurable objective for which resources are allocated; 2) enable reliable information to be independently audited; 3) offer a comparison between different projects based on previous results and costs. One uncovered topic is the effects of EU funding on countries scientific impact. Estonia has been one of the most active participants in EU funding[4] and has increased its scientific impact in terms of citations per paper by 54 per cent during 2007–2014.[5] Therefore, it is important to ask how much of this increase was due to EU funding.

Gains in impact can be explained by the scientist’s credibility cycle. According to García and Sanz-Menéndez,[6] the scientist’s credibility cycle is a relationship between production, communication and collective evaluation of the results, what expands through the process of competing for funding to carry out research. Usually, a resource allocation takes place in a peer review system. In this system, a scientist applies for funding and peers decide on the project funding. In addition to information about the project, peers take into consideration basic aspects about the applicant. One of these aspects is credibility (reputation), which is based on his or her past achievements and affiliations. Also, the same principals apply when scientists are choosing collaboration partners. Therefore, more integrated Estonia gets into the EU science structures and global scientific networks, more possibilities will Estonian
scientists have to improve their impact by having better access to data, resources, equipment and ideas in general.

The enormous growth of collaboration among research institutions and nations worldwide witnessed during the last decades is a function of changes in the dynamics of science as well as science policy initiatives. For researchers from a small country, collaboration may not be only a possibility, but also a necessity to overcome the problems of conducting world-class research in a small country. It can explain why Estonia participates in EU Framework Programmes in large numbers. The smallness of a country is viewed as a constraint for building up domestic human and financial resources for science and expertise in different fields. It is claimed that small countries can compensate such disadvantages in resources through international collaboration. Also, it is found in cases of other EU member states like Spain, Slovenia and the United Kingdom that internationally collaborated papers receive more citations.

Despite positive assumptions that EU funding and international collaboration will increase scientific impact, the relationship may not be so clear. There are some legitimate reasons to believe why international funding and international collaboration may hinder impact in terms of citations (or papers) cost per unit spent. For instance, transaction costs are usually an inevitable outcome of collaboration. In some cases, costs of collaborating may be too large compared to the benefits. For example, travelling and paperwork may take too much time and this could potentially lower the quality of a paper.

The purpose of this article is to find out if EU funding and international collaboration increase Estonian scientific impact. The first hypothesis of this paper is that articles with EU funding have a higher scientific impact than articles with only national funding. The second research hypothesis is that international collaboration articles have a higher impact compared to articles with only Estonian authors. Consequently, our last hypothesis is that EU-funded publications with international collaboration (co-author) produce the best possible outcome in terms of scientific impact.

**MATERIALS AND METHODS**

**Methodology**

This paper provides a quantitative possibility of measurement previously used by Morillo, and complements and applies it to international funding and collaboration. We analyse the set of hypotheses described above by using a database of publications for scientific journal articles that have at least one author from Estonia, published between 2008–2015, included in the Thomson Reuters’ citation indexes. Scientific collaboration usually leads to co-authored papers and publications’ citations in WoS are one the most common measures of scientific impact. The present study takes advantage of WoS search refinement options and InCites possibilities of data gathering. For determining EU and national funding, we use funding acknowledgement (FA) section; included in WoS since 2008. It provides data about the sources of financial support for the research presented in the paper.

We define EU funding if an article in InCities database is marked as funded by one of the following funding agencies: European Union (EU); European Research Council (ERC); European Commission Joint Research Centre; European Science Foundation (ESF); European Cooperation in Science and Technology (COST). For determining national funding, we use Eesti Teadusfond (ETF) and Ministry of Education and Research, Estonia. If an article got funding from national sector and also from the EU, we categorise it as ‘Both’. A publication that has FA but does not belong to any mentioned categories is defined as ‘Other’. This category is too fragmented to bring out precise funding agencies. It mostly contains different foreign funding agencies like Wellcome Trust or the Academy of Finland. Unfortunately, this category may also include EU funding if a funding agency was not in our EU affiliated funding agencies list.

Since the main purpose of this paper is the analysis of the effects of EU funding on scientific impact of Estonia, we drop publications with more than 16 authors. The exclusion of highly collaborated papers restricts our research to the papers that have a substantial contribution from Estonian authors. When the efforts are on a grander scale, with a study group involved, 100 or 50 researchers could not possibly have written, edited and approved the final work. Publications with a high number of co-authors can reflect another type of collaborative effort and not necessarily the ‘actual’ network embeddedness of researchers. Most papers with that many authors are the results of extensive collaborative ‘big science’ projects that conform to a set of procedures and dynamics different from those in smaller groups. In the Estonian case, we are talking about CMS (Compact Muon Solenoid) collaboration what strongly influences the co-authorship geography of small countries. When these publications are not excluded, we will risk getting biased results when evaluating countries with small research systems.

The data on publications include information about citations; date of publication; names and number of authors; address information, funding agencies used; and field of knowledge. For five types of FA— (without FA; national; EU; both; other) different measurement variables were calculated: percentile in the subject area; the portion of articles in the first quartile; the portion of articles with international and domestic collaboration; journal impact factor; and a number of authors. All variables used in the empirical analysis are described in Table 1.
Hirv: Effects of EU Funding on Estonian Scientific Impact

Table 1: List of the Variables Used.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Percentile in the subject area (the main dependent variable)</td>
<td>The percentile in the subject area in which the paper ranks in its category, document type and database year, is based on total citations received by the article.(^{[24]}) The higher the number of citations, the smaller the percentile number. The maximum percentile value is 100, indicating 0 citations received. Because, in a departure from convention, low percentile values mean high citation impact (and vice versa), the percentiles received from InCites are called ‘inverted percentiles’.(^{[24]}) Percentile in the subject area as a measurement of impact (dependent variable) is preferred to category or journal normalised citation impact because it is less sensitive to outliers.</td>
</tr>
<tr>
<td>Funding type</td>
<td>Without FA; National; EU; Both; Other</td>
</tr>
<tr>
<td>Research area</td>
<td>Collected articles were divided into research areas. Research areas were made by regrouping categories from ESI (Essential Science Indicators) scheme into Estonian Research Information System`s four categories: Technology and Engineering; Natural Sciences; Health; Social Sciences.</td>
</tr>
<tr>
<td>The portion of articles in the first quartile</td>
<td>Articles that are in the first quartile of the most cited articles.</td>
</tr>
<tr>
<td>International and domestic collaboration</td>
<td>If there is in addition to Estonia some other country’s address in the article’s address section, we read as a product of international collaboration. If there are two domestic addresses in the address section, we read it as a national collaboration.</td>
</tr>
<tr>
<td>Publication date</td>
<td>Publication date is the date on which a publication is first published. It is included in regression models to take into account possible macroeconomic trends.</td>
</tr>
<tr>
<td>Journal impact factor</td>
<td>Describes how much of an impression a scientific journal makes where an article was published. Shows how many authors (1-16) were involved in writing an article. It is included in regression models to take into account physical labour contributed to an article. Research group characteristics may also impact the funding-productivity nexus.(^{[26]}) Kyvik(^{[27]}) for instance, brings out that larger laboratories may be better positioned to draw together research groups for competitive research grants. They may also have better equipped and be more likely to attract top researchers.(^{[27]}) Therefore, large groups may be better placed to attract and make use of funding which could translate into a higher impact.</td>
</tr>
<tr>
<td>Number of authors</td>
<td>Some publications may collect citations faster than others, although the end result may not differ. This makes an evaluation of recent research publications robust. Also, there is a problem with how funding agencies are determined in WoS.(^{[32]}) For instance, funding source can have several different names and can appear with different conventions for abbreviation, punctuation and form. This can cause a miscategorisation of articles. In addition, the scientific impact may not be the best indicator for smaller countries because their scientific needs may differ compared to larger countries. According to Nygaard,(^{[33]}) scientists have to share their focus on different institutional environments—local, national and international. When topic addresses only local or national environments and ignores the global environment, then a publication is in a disadvantaged position because it excludes a large number of potential readers.</td>
</tr>
</tbody>
</table>

Statistical analyses were applied by means of Stata 14.1 and IBM SPSS 23: comparison of column means (Welch’s \(t\)-test); a decision tree; and Probit combined with truncated OLS regression (with robust standard errors). Approximately one-fifth of the publications in the sample do not have a single citation and these publications are ranked in the 100\(^{th}\) percentile. Truncation is necessary for tackling overrepresentation in this percentile. Year dummies were added to regression to take into account possible macroeconomic trends. In these models, publication percentile in the subject area is the dependent variable and funding type, collaboration and number of authors are explanatory variables. The number of authors squared was added to make models more flexible but is dropped in Probit models because of marginal effects.

Described approach can provide only a part of the information and should be complemented with other approaches to obtain the complete overview of the effects of funding sources and collaboration. The largest problem with this type of funding analysis is the problem of endogeneity. While much attention has been given in the previous quantitative literature to the evaluation of the impact of competitive funding on scientific productivity, modelling issues surrounding endogeneity have remained.\(^{[26]}\) As mentioned previously competitive funding is not allocated exogenously but endogenously determined through prior scientific performance, with funding generally awarded to the ablest researchers. Separating the effects of funding from researchers’ abilities is a complicated problem. In the case of a relationship between funding and publication, publications are usually assigned simultaneously to several researchers. We use research group size (number of authors) as a proxy for principal researcher’s ability as Kyvik\(^{[27]}\) states larger research groups (laboratories) are more likely to attract top researchers. This statement is consistent with previous findings that show there is a causal relationship between the authors involved and impact of the article.\(^{[28,29]}\) Also, Carayol and Matt\(^{[30]}\) argue that analysing funding on a more aggregate level (university or research group) will result in smaller measurement errors. Some other disadvantages of this approach include the time dependence of citations.\(^{[31]}\) Some publications may collect citations faster than others, although the end result may not differ. This makes an evaluation of recent research publications robust. Also, there is a problem with how funding agencies are determined in WoS.\(^{[32]}\) For instance, funding source can have several different names and can appear with different conventions for abbreviation, punctuation and form. This can cause a miscategorisation of articles. In addition, the scientific impact may not be the best indicator for smaller countries because their scientific needs may differ compared to larger countries. According to Nygaard,\(^{[33]}\) scientists have to share their focus on different institutional environments—local, national and international. When topic addresses only local or national environments and ignores the global environment, then a publication is in a disadvantaged position because it excludes a large number of potential readers.
Hirv: Effects of EU Funding on Estonian Scientific Impact

Data

For analysis, research articles with Estonia in the address field in years 2008-2015 were collected. During this period 12445 research articles were published. 1454 articles were excluded because a number of authors exceeded the desired limit. Based on the collected data (Figure 1) we see a steady increase in a total number of articles and also in international collaboration articles.

We see in Figure 2 that after 2008 the proportion of articles without FA has decreased dramatically. It is probable that the decline of 27 percentage points does not present the actual changes in science funding because in later periods we do not see so sudden changes. It is very likely this difference comes from data entering misunderstandings because the FA section was introduced in 2008. For making sure, that does not contaminate our results we exclude publications from 2008 from this point on.

Figure 3 shows that there are substantial differences among research areas in the representation of FA. Natural Sciences has the largest percentages of articles with FA. Taking into consideration that for this period WoS does not fully include FA-s for Social Sciences and it is not surprising that it has very few publications with FA in our sample. Having confirmed the problem, we exclude Social Sciences from the rest of the study.

After the exclusion of Social Sciences, the sample is divided between research areas followingly: 62.3% Natural Sciences; 29.3% Health; and 8.4% Engineering and Technology. 9873 observations stayed in our sample – 72% of observations have a FA(s) and 28% do not.

RESULTS

In general, significant differences between types of funding were found as shown in Table 2. Those do not share a subscript differ at the significance level of 0.05. For example, in Technology and Engineering, articles without FA have an average percentile in the subject area of 66.05. This value is significantly different from articles with national funding (55.41) and from other values in various funding types. Estonia has a remarkably low proportion of articles with national collaboration. This supports Arunachalam and Doss’ argument that in small countries, it is hard to find suitable partners and international collaboration is necessary to overcome this obstacle. Surprisingly, in all research fields, national collaboration is higher when articles are funded simultaneously by national sector and the EU. The probable reason for that is articles with EU funding are a part of larger projects, hence a higher number of authors.

In every research area, the less visible articles are without FA. This is expected because without FA contains articles where only universities’ resources were used. The noticeable differences in scientific impact between nationally funded (National) and EU-funded articles (EU) occur in Natural Sciences and Health. EU-funded articles are ranked in Natural Sciences 11.8 and Health 16.4 inverted percentiles lower (greater impact) compared to nationally funded articles. Also, articles funded by the EU tend to be published in journals with a greater impact factor, are more likely a result of international collaboration and have a higher number of authors. Articles funded
 simultaneously by national sector and the EU (Both) tend to stay in the middle in the mentioned criteria.

For analysing scientific impact in depth, a decision tree was created (Figure 4). A decision tree is formed to explore how funding type and international collaboration affect the percentile in the subject area (dependent variable). For generating the tree, we use CHAID technique and a significance level of 0.05 for splitting and merging decisions. The presented decision tree inclines toward Natural Sciences (the largest research area in the sample) and does not represent Estonian science as a whole because Social Sciences are excluded. As we can see from the decision tree, the best possible combination to maximise research impact is to use EU funding and have an international collaboration partner (co-author). EU-funded internationally collaborated articles (35.83) are ranked in citations ten inverted percentiles lower (higher impact) than nationally funded international collaboration articles (46.26) and seven inverted percentiles lower than articles that got funding from both sources simultaneously (43.08). It suggests that the EU can help Estonia to get a better outcome in international collaboration than otherwise possible. The effect of EU funding is different when we look at articles without international collaboration; the impact does not differ when comparing EU and national funding, but a combination of both simultaneously improves impact significantly.

The results of the regression models are presented in Table 3. Regression Probit<100 reflects the probability that publication has a citation(s). Truncated OLS gives effects on the dependent variable in percentiles given that publication has a citation(s) and Probit Q1 model reflects the probability that publication is in the upper 25th percentile of the most cited articles. Although there is some variability, regression models support a decision tree finding that combination of EU funding and international collaboration produces the most cited scientific articles. This is especially seen in Natural Sciences but also in Health. Surprisingly, in Technology and Engineering, articles funded simultaneously by the EU and the national sector have nine inverted percentiles less impact than nationally funded articles and international collaboration is not a significant factor at all when controlling for a number of authors involved.

In Health, international collaboration improves article’s probability of having a citation(s) by 11.9 percentage points. When looking articles with at least one citation, then international collaboration lowers their rankings (higher impact) by 8.4 inverted percentiles. Also, it improves the probability that an article is in the first quartile by 12.3 percentage points. In addition to collaboration, EU funding also has also a positive effect. EU-funded publications are ranked seven inverted percentiles lower and are also more likely in the first quartile than nationally funded ones. We see very similar results in

![Figure 4: Decision tree of the percentile in the subject area by funding types and international collaboration (WoS 2009–2015).](https://example.com/fig4.png)
Natural Sciences. One noticeable difference between these research areas is that in Natural Sciences also articles funded simultaneously by the EU and national sector have a significantly higher impact than only nationally funded articles. The positive effect of international collaboration is in Natural Sciences in the same range as in Health.

Also, national collaboration has some effect on impact. In Health it improves articles’ probability of getting cited and in Natural Sciences it improves impact by two inverted percentages given that publications have a citation(s). It is an example that in some cases smallness is beneficial for national collaboration because more flexible and transparent institutional system of research will generate higher density and frequency of relationships (know-who)\cite{29,28} what can lead to increased scientific impact.

A number of authors is a significant factor of research impact in all research areas. It seems that a larger team is necessary for writing a well-cited article, although in Health it is not so important as in other fields. The effect of the number of authors involved is nonlinear in Technology and Engineering (Figure 5) and supports Kenna and Berche\cite{29} argument in this matter – an increase in authors involved improves research impact only to a certain degree. Once critical mass is attained, a research team has used its opportunities for cooperation as well as improving access to more resources.\cite{29,28} In Technology and Engineering, the optimal research group size is around ten.

Wald’s test whether time dummies are jointly zero is in most of the models insignificant. In most of the cases, publication date does not have an effect on normalised publication impact. Although publication date is significant in some Probit models (Probit <100) where we test whether an article is cited, but this is explained by articles’ time in circulation. More time publication has been available; the higher is the probability that paper has a citation(s). Time dummies are jointly significant.
Firstly, EU funding provides a higher research impact than national funding. This was seen especially in *Natural Sciences*, where articles funded simultaneously by the EU and national sector were significantly more cited than only nationally funded ones. Secondly, international collaboration improves scientific impact significantly. Thirdly, the results show that a combination of EU funding and international collaboration produces the most cited scientific articles. These results suggest that regarding scientific impact, the EU can help Estonian scientists get a better outcome in international collaboration than otherwise possible.

The positive effect of EU funding was not seen in *Technology and Engineering*. In this research area, articles funded simultaneously by national sector and the EU had less impact than nationally funded articles when controlling for a number of authors involved. Also, in this research area, international collaboration was not a significant factor determining scientific impact. A possible reason may be that this research area has been developed as a national priority and therefore there has not been an incentive to have strong international partners. Also, a very probable reason is that conference proceedings and articles/chapters in books are not included in the analysis and this may have an impact on results in this research area.

The main limitations of this paper include possible problems with endogeneity, the time dependence of citations what makes an evaluation of recent publications robust and methodical problems how funding agencies are determined in WoS. For further studies, it is necessary to determine the largest partners of Estonian scientists and their impact. For example, how strong are Estonian research partners in *Technology and Engineering*? Also, it is unknown what factors (capital; know-how; or both) do Estonian scientists get from EU funding and international collaboration or at least to some extent they make scientists ignore national institutional environments and focus on international environments where topics call for a greater number of readers.

**ACKNOWLEDGEMENT**

This article is based on the applied research funded by Estonian Ministry of Education and Research, ERDF and Estonian Research Council via RITA programme.

**CONFLICT OF INTEREST**

The authors declare no conflict of interest.

**ABBREVIATIONS**

EU: European Union; WoS: Web of Science; FA: Funding acknowledgement; CMS: Compact Muon Solenoid; ESI: Essential Science Indicators; ERC: European Research Council; EC: European Community;
REFERENCES