Citations:

Citations were by Scopus for 11 Units of Assessment (<http://www.ref.ac.uk/2014/about/guidance/citationdata/>), with the following ones included in the selected sample:

* Public Health, Health Services and Primary Care (Panel A)
* Biological Sciences (Panel A)
* Physics (Panel B)
* Computer Science and Informatics (Panel B)
* Economics and Econometrics (Panel C)

According to the panel criteria and working methods, as supplied by the panel, ([http://www.ref.ac.uk/2014/pubs/2012-01/#d.en.69569](http://www.ref.ac.uk/2014/pubs/2012-01/" \l "d.en.69569) ) the citation counts for Panel A, Panel B and Panel C where used as a supplementary tool as part of the indication of academic significance of the published work. However when reviewing the published criteria, Panels A and B appeared to have given more emphasis on the citation count than Panel C, with Panel B even utilising additional citation data from Google Scholar as an indicator. Panel B criteria (<http://www.ref.ac.uk/2014/media/ref/content/pub/panelcriteriaandworkingmethods/01_12_2B.pdf>) have specified that Google Scholar was accessed in a systematic way to observe if there was work highly cited outside of Scopus, which raises some issues. It does not specify if the Google Scholar citation data was retrieved during the same period for all papers and to what extent was this data used. Nevertheless , what one can interpret from the published criteria from these panels, is that the citation data should partly be responsible for the output assessment.

In the panel overview reports which were conducted after the REF 2014 process was finalised, Panel B results state that the citation data did not have any effect on the outcome of the ouput results.

#### Possibly the following section can go in the Litterature Review?####

Previous work on how citations could have influence the outcome of the REF 2014 scores have been performed by Alan Dix (<http://alandix.com/docs/ref2014/bib-2015-citations-and-ref-v2.pdf>) and Morris Sloman (<http://alandix.com/docs/ref2014/SP11-REF-analysis.pdf> ). Morris Sloman focused on analysing the results from the Computer Science and Informatics UOA. What Sloman has concluded in his work, is that there is a correlation between the citations and the REF output. Alan Dix’s work “Citations and Sub-Area Bias in the UK Research Assessment Process”, also focused in the same UOA, and what he has concluded from his analyses is that there is bias between the sub-areas of the UOA.

##### Hypothesis

The hypothesis formed after reviewing the information above, is that the number of citations affects the output percentage of a university. The expected results are that with increasing citation numbers there will be an increase in 4\* papers and a decrease in 2\* and 1\* papers.

##### Methodology

The analysis was performed only on the UOAs listed above, due to the confirmed source of the citation data (<http://www.ref.ac.uk/2014/about/guidance/citationdata/>), using the data provided by the REF (<http://results.ref.ac.uk/DownloadSubmissions/SelectUoa> ). The original data contained the citation counts as integers, including 0, for each submission and in certain cases data was not available and the citation count was given as blank. To account for the unavailable data, instead of comparing the output score against the total citations per university, the total citations were divided by the number of submissions in each university to obtain a normalised value.

The citations per submission for each university were then compared against the percentage of output raked 4\* and the combination of 2\* and 1\* papers in each university. The Pearson’s method was used to return the correlation of the two variables.

Figure.. shows the plotted results for the Economics and Econometrics and Biological Sciences UOAs. From the figures it is observed that the results produced against 4\* and the average citations per submission in each university, have more of a linear relationship. When comparing against the combined 2\* and 1\* outputs the relationship in the Economics and Econometrics UOA appears to follow a quadratic pattern. Were this was true Spearman’s method was also used, which is a non-parametric version of the Pearsons R test (<https://statistics.laerd.com/statistical-guides/spearmans-rank-order-correlation-statistical-guide.php>[),](https://statistics.laerd.com/statistical-guides/spearmans-rank-order-correlation-statistical-guide.php),and) and the results were almost identical and did not affect the outcome.

### Insert Figure...The following table shows the results obtained in the UOA’s tested.

|  |  |  |
| --- | --- | --- |
|  | **Citations per submission against:** | |
| **UOA** | **Output 4\*** | **Output 2\* + 1\*** |
| **Public Health, Health Services and Primary Care** | 0.63 | -0.52 |
| **Biological Sciences** | 0.40 | -0.35 |
| **Physics** | 0.49 | -0.61 |
| **Computer Science and Informatics** | 0.57 | -0.63 |
| **Economics and Econometrics** | 0.73 | -0.63 |

The results suggest that there is large variance between the effect of the citation numbers and output scores in each UOA. In the Economics and Econometrics UOA it appears that the assessment of the output 4\* scores was influenced by the citation counts, as the r value returned indicates a strong positive correlation. The rest of the results, excluding the Biological Sciences UOA, suggest that there is moderate positive correlation between 4\* output and the citations per submission per university and a moderate negative correlation between the combined 2\* and 1\* output per university and the citations per submission per university. The results returned for the Biological Sciences can be interpreted as weak correlations and therefore no useful results can be extracted.

Income:

While Sloman in his work (<http://alandix.com/docs/ref2014/SP11-REF-analysis.pdf> ) compared the Income/ FTE against the environment output of the university, in this case the average income per submission for each university was compared against 4\* output and also against the combined 2\* and 1\* output.

The hypothesis under this feature is that the research income of each university affects the output results. The expected results are that higher incomes will indicate a higher percentage of 4\* outputs and a lower percentage of combined 2\* plus 1\* outputs.

The original data contained the income per source per year for each university. To obtain the average income per submission, the total income over the five year period was summed up and divided by the total number of submissions for each university. To assess if there is a relationship between the variables, the Pearson R correlation test was used. The results returned can be seen in Table (table below).

|  |  |  |
| --- | --- | --- |
|  | **Income per submission against:** | |
| **UOA** | **Output 4\*** | **Output 2\* + 1\*** |
| **Public Health, Health Services and Primary Care** | 0.45 | -0.39 |
| **Biological Sciences** | 0.82 | -0.70 |
| **Physics** | 0.45 | -0.48 |
| **Computer Science and Informatics** | 0.74 | -0.67 |
| **Aeronautical, Mechanical, Chemical and Manufacturing Engineering** | 0.65 | -0.52 |
| **Economics and Econometrics** | 0.73 | -0.60 |
| **Anthropology and Development Studies** | 0.38 | -0.47 |
| **Education** | 0.41 | -0.36 |
| **Philosophy** | 0.63 | -0.55 |
| **36 - Communication, Cultural and Media Studies, Library and Information Management** | 0.34 | -0.22 |

The results indicate that there is a strong correlation between the 4\* output and the income per submission in the Biological Sciences, Computer Science and Informatics and Economics and Econometrics. This can suggest that possibly the research quality in these UOA’s is heavily influenced by the average research income per submission.

Journal Rank:

The hypothesis is that the journals in which the research work was submitted to affects the 4\* output score of the papers. To assess this hypothesis the journal list relating to the 2010 Excellence in Research for Australia (<https://www.righttoknow.org.au/request/journal_list_relating_to_the_201> ) was used, which ranks journals by A\*, A, B, C and Not Ranked, with A\* being the highest rated.

The original data from the REF contains (per UOA) contains the submission, volume title and the ISSN code given for the volume. Since there is a four year gap between the publication of the REF and the 2010 ERA, it is assumed that the journal ranks remained unaffected. Additionally due to the previous fact and the fact that not all work submitted in the REF was published in journals, it was expected that there would be some data loss when matching the two data sets.

The two data sets were initially matched by volume titles, with most UOA’s having half of the data matched. To match by volume titles, the text was first lower-cased, punctuation was removed and any symbols. This raised the issue though that there might be a mismatch between similarly named journals. To avoid this it was decided that it would be better if the datasets were matched by the ISSN numbers.

The matched dataset contained the university, submission, volume title and rank, were matched with the 2010 ERA. To be able to quantify the effect the journal ranks have on the output, the journal rank scores were given a representative score. The scores were converted as follows:

A\* - 4

A – 3

B – 2

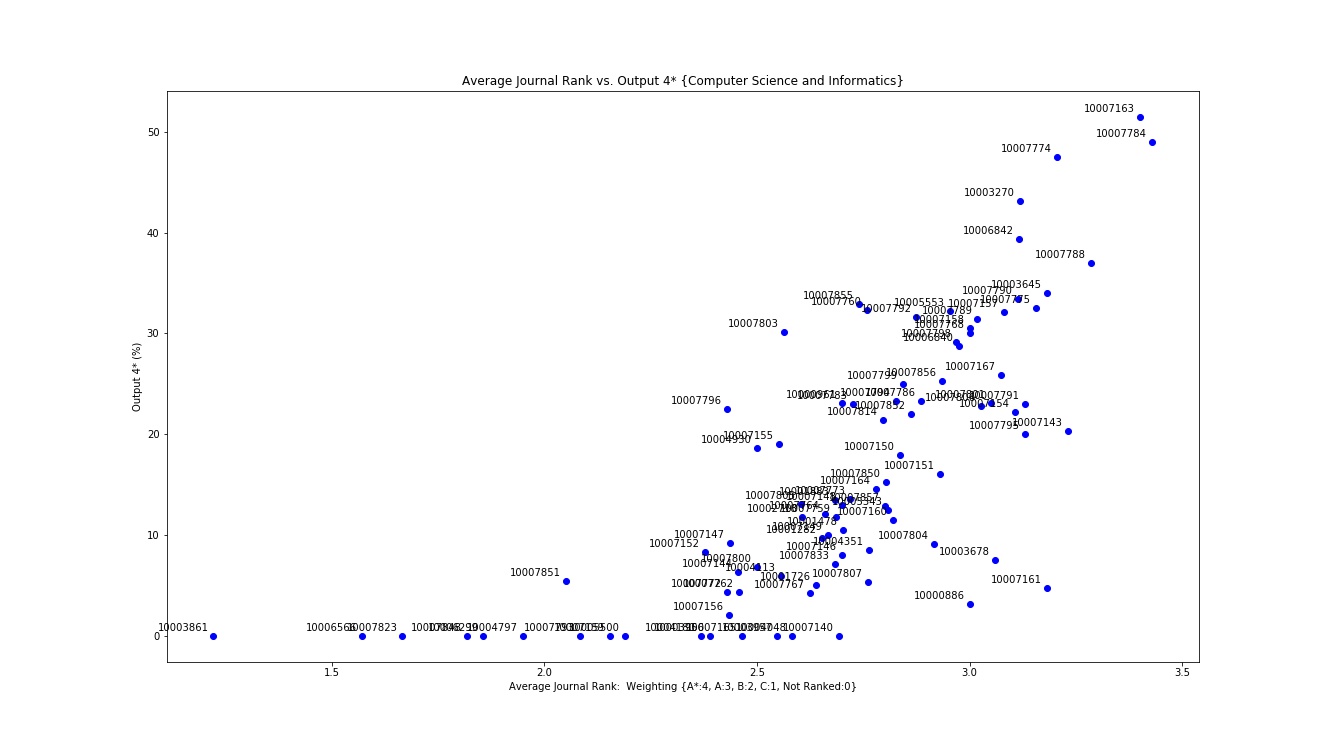
C – 1

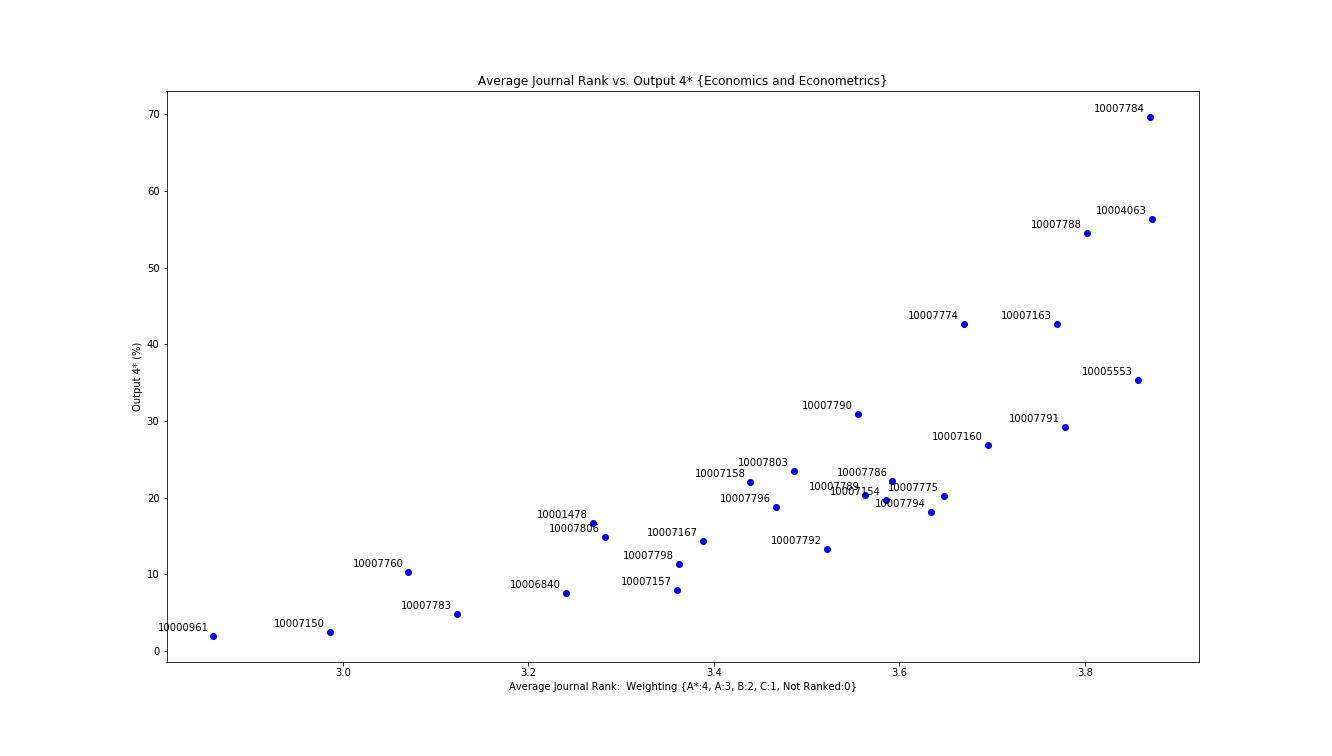
Not Ranked – 0

The idea behind the scores is that a paper published in an A\* journal will have a larger gravity in affecting the output 4\* percentage of the university, C will have the least and Not Ranked will have none. Different weighting schemes were trialled on several UOAs before settling with the scores above. Different scores produced better correlation in some cases, for example in the Physics UOA, the scores A\*: 1, A : 1, B : -2.5, C: -5 and Not Ranked:0, returned a correlation of 0.61. This is better but the pattern was the same and the overall outcome was not largely affected. After converting the 2010 ERA ranks to their scores, the scores were summed up for each university and divided by the number of submissions to find the average journal rank per university. The average journal rank per university was then compared against the 4\* output. The results returned for each UOA can be seen in Table … (Table below).

|  |  |  |
| --- | --- | --- |
| **UOA** | **Matched Data (%)** | **Output 4\*** |
| **Public Health, Health Services and Primary Care** | 90% | 0.63 |
| **Biological Sciences** | 95% | 0.80 |
| **Physics** | 92% | 0.53 |
| **Computer Science and Informatics** | 62% | 0.73 |
| **Aeronautical, Mechanical, Chemical and Manufacturing Engineering** | 92% | 0.55 |
| **Economics and Econometrics** | 89% | 0.82 |
| **Anthropology and Development Studies** | 60% | 0.57 |
| **Education** | 53% | 0.43 |
| **Philosophy** | 54% | 0.60 |
| **36 - Communication, Cultural and Media Studies, Library and Information Management** | No results could be extracted | |

The results from several UOAs returned a strong positive correlation, implying that the journal rank does affect the output score of a paper. In an attempt to further analyse the results the Computer Science and Informatics UOA was selected. The figure below shows the plotted results for the UOA.

What is visible from the graph is that almost all of the universities with up to approximately 2.3 in average journal rank have 0% of 4\* output rated papers. After that threshold the data appears to follow a linear pattern, with universities having a higher average journal rank having also a higher percentage of 4\* output. What is interesting are the universities that appear not to follow the pattern, or are no close to the rest of the data. Specifically universities with UKPRN 10000886 and 10007161 have a high average journal rank but a very low percentage of 4\* output papers, when comparing with universities with similar average journal rank. What was observed is that these universities have significantly less submissions than the universities with an average rank close to these two. To test this theory further the UOA of Economics and Econometrics was also selected for further analysis. By selecting the university with the UKPRN 10005553 and comparing to the universities with a higher 4\* output but similar journal rank, the same pattern is observed. University 10005553 has significantly less submissions than the universities with a close average journal rank. The complete tables of results for the two UOAs selected for the further analysis, can be observed in Appendix..



To futher test this hypothesis the percentages of papers published in A\*, A, B, C and Not Ranked journals for each university were gathered and all tested individually to all the output scores. The results returned suggested there is little or no correlation between all the tested combinations and therefore these were not explored any futher.

Therefore the analysis on the journal ranking feature suggests that the higher the average journal rank in the UOAs with strong correlation the higher the possibility of a larger percentage of 4\* output score.

SJR Score vs. 4\*:

To further assess the strongly correlated UOA’s and strengthen the hypothesis, journal assessment scores from the SCImago Journal & Country Rank ([http://www.scimagojr.com](http://www.scimagojr.com/aboutus.php)) for the year 2014 were retrieved, matched with the REF data to obtain an average journal score per university against the 4\* output. The data was matched by the journal name, because unlike with the 2010 ERA data the ISSN matched very low percentages. The matched data for the three UOA’s selected for the analysis reached 70%. Their indicator, the SJR score was used, which is based on data from the Scopus data. The UOA’s selected for this analysis are: Biological Sciences, Computer Science and Informatics, Economics and Econometrics. The table below summarises the results of the analysis.

|  |  |
| --- | --- |
| **UOA** | **Output 4\*** |
| **Biological Sciences** | 0.83 |
| **Computer Science and Informatics** | 0.63 |
| **Economics and Econometrics** | 0.95 |

What is interesting and we can observe in the figure below is that the Economics and Econometrics UOA shows a very strong correlation between the average SJR score per university and the output 4\* .

