**Authors’ response**

Dear Professor Ruiz,

Thank you very much for your valuable comments. My co-authors and I have made the changes suggested by you and the following are our responses to the specific points that need change as per your comments.

***Comment 1:***

*The idea of doing it in Rmarkdown and in a git repository was to make it* *reproducible. But all the analysis code has been omitted, so it is not reproducible at all. My idea was to put the data and code in the repo, so that when the data is updated, we can easily re-run the code. If the analysis is in something other than R, that doesn't matter, as it can be included separately in the repo. The fundamental principle is reproducibility. We need to be able to update the data from scopus, and regenerate all tables and graphs by simply recompiling the Rmarkdown file (and possibly an additional script file).*

***Our response to comment 1:***

Thank you for your comments. First, we need to declare that all of the tables and figures in the manuscript are reproducible. The retrieval strategy is Source Title= International Journal of Forecasting AND Timespan= 1985-2018. The retrieve date is on March 25th. The raw data including 1938 IJF papers is harvested from Scopus. Anyone can derive the same dataset based on this retrieval strategy in the Scopus database, but the number of citations of each paper may vary due to different retrieve date.

All of the tables can be reproduced as below:

As for the table 1, data cleaning is conducted first. Papers that lack of the affiliation information are excluded. Then the top 20 most cited papers are identified based on the number of citations of each paper. Three indicators: Country, Nau (the number of authors), and Nin (the number of institutions) are calculated based on the affiliation information.

As for the table 2, the raw data is imported into bibliometric software to derive a citation network, then the papers that are embedded in the citation network can be identified. After identifying these papers, their number of citations and local citations (the citations from the IJF papers) are calculated. The number of citations of each paper can be obtained directly from the raw data. The number of local citations of each paper can be calculated based on its citation information.

As for the table 3, the raw data is imported into the bibliometric software to derive a co-citation network, then the top 10 co-cited papers are identified based on the co-citation number of each paper. Besides, a citation history function in the bibliometric software can be used to derive the values of the three indicators which are FY (The first year being co-cited), LY (The last year being co-cited), and Most co-cited year (number).

As for the table 4, the raw data is imported into the bibliometric software, and a burst detection function is used to derive the top ten papers with the strongest power in the burst detection. The values of the three indicators can be derived from the burst detection function, and these three indicators are Strength (the value of burst detection), Begin (the first year being cited), and End (the last year being cited).

As for the table 5, data cleaning is conducted first. Papers that lack of the affiliation information and author information are excluded. Then the top ten prolific authors are identified based on the number of papers he/she published. The indicator of Country can be processed based on the affiliation information. The indicators of 1st (the number of papers whose author is as the first author) and Cor (the number of papers whose author is as the corresponding author) can be processed based on the author information.

As for the table 6, the top ten most cited authors are identified based on the number of citations that his/her papers received. The indicators of TP (the number of papers the author published) and LC (the citations from the IJF papers) can be calculated based on the citation information in the Scopus database.

As for the table 7, the top ten prolific countries can be identified in the Scopus database. The values of the indicators TP (the number of papers the country published), FY (The year that the first paper published in the country), and LY (The year that the last paper published in the country) can be derived in the Scopus database.

As for the table 8, the top ten most cited countries can be identified in the Scopus database. The values of the indicators TP (the number of papers the country published) and TC (the number of citations the country published) can be derived in the Scopus database. The values of the indicators In-TC (The number of citations from IJF) and Out-TC (The number of citations outside IJF) can be calculated based on the citation information.

As for the table 9, the raw data is imported into the bibliometric software and then a country collaboration network can be derived. In our manuscript, 25 countries that published no less 15 times are extracted to construct the country collaboration network. The values of the three indicators can be derived from the bibliometric software. These three indicators are Link (the collaboration times a country possesses), CM (the country that collaborate most with the target country), and Number (the collaboration times that CM collaborates with the target country).

As for the table 10, the top ten most prolific subject areas belonging to 29464 citing papers of IJF and 16419 citing papers of JF can be derived in the Scopus database. Therefore, this table can be reproduced. Note that the number of citing papers of IJF/JF may vary over time, so the number of papers belonging to the top ten most prolific subject areas may also change accordingly.

As for the table 11, 1938 IJF papers and their citing papers belonging to Computer Science are imported into the bibliometric software to derive an IJF journal citation network in Computer Science. Then the top ten journals in the IJF journal citation network in Computer Science can be derived in the bibliometric software. Two indicators can be calculated in the bibliometric software, which are Links (the number of citations that journal has within selected IJF citing papers in Computer Science), and Links with IJF (the number of citations that are between the journal and IJF).

As for the table 12-30, the similar operations are processed as the table 11. Therefore, these tables can be reproducible.

All of the figures can be reproduced as below:

As for the figure 1, two indicators (the number of publications annually and the number of citations annually) can be derived based on the raw data. Note that the number of citations annually may change over time.

As for the figure 2, the raw data is imported into the bibliometric software, and the number of citations no less than 80 is set as the limitation, then 111 qualified IJF papers are derived. 29 of the 111 papers are discarded because they are isolated. Finally, 82 IJF papers are used to construct this citation map.

As for the figure 3, the raw data is imported into the bibliometric software, and each group of thresholding has three criterion which are the minimum citations, minimum co-citations, and minimum normalized co-citations. The first thresholding is set in the year of 1985, the second in 2002, and the last is in 2018. After some trail runs, the first thresholding is set with 50, 3, 15, the second is 3, 3, 20, and the third is 3, 3, 20. Finally, the figure 3 is obtained based on these criterion.

As for the figure 4, the raw data is imported into the bibliometric software, and authors who publish less than eight papers are discarded, then 38 qualified authors are derived. Among the 38 authors, 6 authors are isolated, so finally 32 authors are mapped into the co-author network.

As for the figure 5, two indicators are used: percent 1 (the ratio of the yearly IJF papers produced by certain country to the yearly number of IJF papers), percent 2 (the ratio of the yearly IJF papers produced by certain country to the total number of IJF papers produced by certain country). The number of yearly IJF papers produced by certain country, and the number of yearly IJF papers can be derived from the raw data.

As for the figure 6, the raw data is imported in the bibliometric software, and 25 countries that published no less 15 times are extracted to construct the country collaboration network. Besides, a .net file of the country collaboration network derived from the bibliometric software, and we can manually edit the .net file to prevent over-plotting issues. Also the case in names of countries can be changed through editing the .net file.

As for figure 7, the IJF papers and the IJF citing papers belonging to Computer Science are imported into the bibliometric software. 21 journals that have no less than 40 publications and 200 citations are selected to map the IJF Journal citation network in Computer Science.

As for the figure 8-26, the similar operations are conducted to map the Journal citation networks. Different networks have different parameters (the minimum number of publications and the minimum number of citations), and we state the parameters in the table as below:

|  |  |  |  |
| --- | --- | --- | --- |
| Figure name | Network name | Minimum publications | Minimum citations |
| Figure 8 | IJF Journal citation network in Economics, Econometrics and Finance | 50 | 700 |
| Figure 9 | IJF Journal citation network in Business, Management and Accounting | 35 | 400 |
| Figure 10 | JF Journal citation network in Business, Management and Accounting | 20 | 300 |
| Figure 11 | JF Journal citation network in Economics, Econometrics and Finance | 40 | 500 |
| Figure 12 | JF Journal citation network in Mathematics | 20 | 300 |
| Figure 13 | IJF Journal citation network in Engineering | 30 | 300 |
| Figure 14 | IJF Journal citation network in Mathematics | 30 | 300 |
| Figure 15 | IJF Journal citation network in Social Sciences | 20 | 300 |
| Figure 16 | IJF Journal citation network in Decision Sciences | 30 | 300 |
| Figure 17 | IJF Journal citation network in Energy | 15 | 200 |
| Figure 18 | IJF Journal citation network in Environmental Science | 15 | 180 |
| Figure 19 | IJF Journal citation network in Earth and Planetary Sciences | 7 | 100 |
| Figure 20 | JF Journal citation network in Computer Science | 15 | 100 |
| Figure 21 | JF Journal citation network in Decision Sciences | 20 | 300 |
| Figure 22 | JF Journal citation network in Engineering | 12 | 180 |
| Figure 23 | JF Journal citation network in Social Sciences | 10 | 150 |
| Figure 24 | JF Journal citation network in Environmental Science | 10 | 150 |
| Figure 25 | JF Journal citation network in Energy | 5 | 30 |
| Figure 26 | JF Journal citation network in Earth and Planetary Sciences | 4 | 70 |

Based on the parameters stated above, the networks in figure 8-26 can be derived from the bibliometric software. Besides, through editing .net file of each network manually, we can optimize the networks to prevent the over-plotting issue.

Above all, all of the tables and figures in the manuscript are reproducible. According to your comments, we need to put the data and code in the repo, and when the data is updated, we can easily re-run the code. However, we’d like to say that we are sorry about this technical problem. We do make an effort to this technical problem, but this problem still cannot be done 100 percent. We state the primary issues during the process of figuring this technical problem as below:

(1). In the manuscript, the raw data are harvested from Scopus database, and all of the tables and figures are made from the raw data. However, the most important point is that all the tables and figures need complex and trivial data preprocessing. After the data preprocessing, we can use various bibliometric indicators, measures, and software to make tables and figures. Therefore, what we want to emphasize is that the raw data usually needs to be preprocessed manually.

(2). All of the data used in the tables need be preprocessed manually in advance, and most of them are processed in Excel. Due to some frequently-appeared format errors and data missing problems, bibliometric knowledge is necessary for dealing with unpredictable problems during this process. But if all of the preprocessing work is done in an R script rather than manually in Excel, we’d like to say that our bibliometric knowledge cannot be applied flexibly. Without of the bibliometric knowledge, the preprocessing would be a tremendous amount of work, and the cost of this work has already surpassed the value of it.

(3). All of the figures need a secondary operation manually. After importing data and constructing networks in the software, we need to adjust networks manually to derive a better graph composition. Different networks need different adjustments, and we cannot predict what adjustments are needed until the networks are constructed, so it is very hard to do in an R script.

(4). We do make an effort to write an R script, but in the actual operation process, we find out that writing R scripts in the chunk of the RMD file is different from directly writing R scripts. Some R packages can be directly used in R scripts, but fail into the R scripts in the chunk of the RMD file. For example, the “flextable” package is a useful R package for generating tables from data files. The “flextable” package can be used successfully in an R script, but it doesn’t work when we use it in the R script in the chunk of the RMD file. We conjecture it is due to the compatibility problem between RMD file, tex file, and R script.

Above all, although the technical problem in your comments are unsettled, all of the tables and figures in this manuscript are reproducible. Anyone can reproduce our tables and figures with the raw data derived through our retrieval strategy.

***Comment 2:***

*The tables in the Rmd file are hard-coded. They should be generated from the data using scripts.*

***Our response to comment 2:***

Thank you for your valuable comment. The tables in the RMD file are hard-coded. Therefore, in this revision, we delete all of the table code in the RMD file, and write the code into the “Preamble. tex” file. Then we use a newcommand code in the “Preamble. tex” file to define a new table function, and use this new table function in the RMD file to generate the tables. Through this operation, we can derive a RMD file without hard-coded tables, and we hope it can be helpful to your comments.

In this revision, we do try to solve this problem as you mentioned. We save the data into excel files (both of the csv and xls formats are tried), and then try to generate tables from the excel files using R scripts. However, during this process, we find a question. It is easy to derive the tables in a PDF file, if we directly program an R script to generate tables from the excel files based on using an R package “flextable”. But it doesn’t work in the RMD file. We cannot use the “flextable” package in an R script chunk in the RMD file, which we think is because of the compatibility between the RMD file, latex file, and R script. We have searched for a long time but cannot find any solution about this compatibility issue.

We are very sorry about this unsettled compatibility issue, but we do alleviate the problem that the tables in the RMD file are hard-coded. Please check it in the RMD file and “Preamble. tex” file.

***Comment 3:***

*The eps figures are all very* *low res, and not vector-based. The main advantage of eps is that you can create vector-based graphics which are crisp and* *scalable. I don't know how you are producing the figures, but it should be possible to save the results as* *vector-based eps or pdf figures.*

***Our response to comment 3:***

Thank you for your helpful advices. In the previous manuscript, all of the journal network graphs were produced by a bibliometric software, and they only could be exported as PNG graphs. Then PNG graphs were transferred into EPS graphs, and inserted into RMD file successfully. Therefore, we actually do not have vector-based EPS graphs because we just obtain the EPS graphs though transferring from the PNG graphs. The low res of EPS graphs is due to the low res of PNG graphs.

In order to solve the problem of the low res graphs, in this revision, we use another image processing software, Photoshop, to transfer these low res PNG graphs into high res PNG graphs. Then transfer the high PNG graphs into EPS graphs, and insert EPS graphs into the RMD file. We are sorry about the problem you mentioned. We cannot change the format of the graphs produced from the bibliometric software. What we can do only is to increase the pixels of the PNG graphs as much as possible. In this revision, all of the graphs in the manuscript have been processed again to increase the pixels as much as possible. Please check it.

***Comment 4:***

*Figure 1 does not need 5 panels. Just give the main figure at the same width as the page.*

***Our response to comment 4:***

Thank you for your comment. We have re-made the figure 1, and just give the main figure at the same width as the page. Besides, the content related to the figure 1 in the **Section 2.1** (Basic statistics of IJF) has been modified. Please check it.

***Comment 5:***

*This figure also appears to involve some smoothing, or bezier* *interpolation which is inappropriate. We should just give the data with linear interpolation.*

***Our response to comment 5:***

Thanks to your thoughtful advices. We have made the figure 1 with linear chart, and given the data with linear interpolation in **Section 2.1** (Basic statistics of IJF). Please check it.

***Comment 6:***

*The number of citations for 2018 is extremely small, possibly because not all 2018 papers are included. We should be clearer about what is included, and what is going on at the end.*

***Our response to comment 6:***

Thank you for your comments. In our previous manuscript, we used the retrieval strategy: Source Title= International Journal of Forecasting AND Timespan= 1985-2018, and the retrieve date is on March 25th. 71 publications published in 2018 have been identified based on this retrieval strategy, and according to the raw data, the publications in 2018 have been cited by 105 times up to March 25th. Therefore, we can assure that all of the publications published in 2018 have been included in our previous manuscript. We think that the number of citations for 2018 publications is extremely small is largely due to the short citation time window. Publications in 2018 do not have enough time to receive and accumulate citations, which leads the number of citations for publications in 2018 becomes small.

Besides, we check the number of publications in 2018 again in Scopus. The same retrieval strategy is used for retrieving data, and the retrieve date is on July 24th. At the end, we derive 71 publications and 174 citations. From March 25th to July 24th, it is reasonable that another 69 citations for 2018 publications have been accumulated. Based on this retrieve result, we can also prove that the number of citations in our previous manuscript is correct. Therefore, we think that the number of citations for 2018 publications is extremely small is largely due to the short citation time window. The data in the figure 1 remains unchanged in order to be in accordance with the following analyses.

***Comment 7:***

*The peak in 2006 is due to a special issue containing largely review papers, looking at the 25 years since the journals and institute began.*

***Our response to comment 7:***

Thank you for your valuable comments. We have identified the special issue published in 2006, and highlighted four highly cited review papers. The related content has been re-written in the first paragraph of the **Section 2.1** (Basic statistics of IJF). The revised content is stated as below. Please check it.

“The number of citations in the year of 2006 is the highest, which largely owes to a special issue containing largely review papers which aims at looking at the 25 years since the journals and institute began. Among this special issue, the top four highly cited review papers are identified: the work of De Gooijer and Hyndman (2006), Gardner (2006), Lawrence et al. (2006), and Booth (2006).”

***Comment 8:***

*There is inconsistency in Figures 2 and 3 with Fig 2 in lower case and Fig 3 in upper case. Also punctuation differences. Perhaps it would be cleaner to use single letters for each paper instead.*

***Our response to comment 8:***

Thank you for your comments. We have unified the figure 2 and 3 in upper case, and deleted all of the punctuations in figure 2 and 3. Please check it.

***Comment 9:***

*The use of a green and red color scheme in Fig 3 won't work for color-blind people. Another way to do it is to use a rainbow palette which effectively divides the papers into about 6 colour groups.*

***Our response to comment 9:***

Thank you for your helpful advices. We have checked the software we used to produce the co-citation network, it does provide a similar drawing function, but the function is more inclined to provide a gradient color rather than 6 color groups. We think that the figure 3 with a gradient color also can work for color-blind people, so we have re-made the figure 3 with a gradient color. The dark color represents the papers published in early years, while the light color represent the papers published in late years. Please check it.

***Comment 10:***

*Inconsistent case in names of journals and names of people.*

***Our response to comment 10:***

Thank you for your helpful comment. We have re-made all of the figures with the unified case in names of journals and names of people. Please check it.

***Comment 11:***

*I don't think we should exclude associate editors. In any case, almost all of the people in Table 4 have been on the editorial board at some point. Anyone who is a top researcher in forecasting will be invited to the board, so the* *causality is not necessarily in the direction you assume.*

***Our response to comment 11:***

Thank you for your helpful comments. We have redone the **Section 2.3** (Authors analysis in IJF and co-author network) based on considering all the prolific associate editors. The **Section 2.3** (Authors analysis in IJF and co-author network) including the table 5, the table 6, the figure 4, and the related content have all been revised. Please check it.

***Comment 12:***

*Why isn't Zhang listed in Table 5? He authored the most cited paper with over 2000 citations so far.*

***Our response to comment 12:***

Thank you for your advice. In the previous manuscript, Zhang is excluded because he only published one IJF paper. In this manuscript, we have remade the table 5 through calculating all of the highly cited authors, regardless of how many IJF papers he/she has published. Note that in this revised manuscript, the table 5 is changed to the table 6. Please check it.

***Comment 13:***

*Fig 5 is hard to read due to the faint names on a dark background. There is no need to obscure names here.*

***Our response to comment 13:***

Thank you for your advice. We have remade the figure 5 with no obscure names. Note that in this revised manuscript, the figure 5 is changed to the figure 4. Please check it.

***Comment 14:***

*Fig 6. Replace these two graphs with a* *time plot for each country. Perhaps just the top five countries to prevent* *overplotting issues.*

***Our response to comment 14:***

Thank you for your advices. We have replaced the two graphs in figure 6 with one graph with a time plot for each country. Note that in this revised manuscript, the figure 6 is changed to the figure 5. Double axes graph is used to describe the top productive countries. Primary axis describes the ratio of the yearly IJF papers produced by certain country to the yearly number of total IJF papers. Secondary axis describes the ratio of the yearly IJF papers produced by certain country to the total number of IJF papers produced by certain country. After several trail run, we find out that top three are the maximum number to prevent over-plotting issues. Therefore, the figure 6 only includes the information about the top three prolific countries. Please check it.

***Comment 15:***

*Fig 7 Some country names obscured.*

***Our response to comment 15:***

Thank you for your advices. We have remade the figure 7 with no obscure country name. Note that in this revised manuscript, the figure 7 is changed to the figure 6. Please check it.

***Comment 16:***

*Section 3. Again, time series plots rather than arbitrary splitting the data into 3 periods would be better. It is reasonable to give only the 3 main subject areas and provide the other information as an appendix.*

***Our response to comment 16:***

Thank you for your valuable comments. We have used time series to replace splitting the data into 3 periods. The **Section 3.1.2** (Dynamic analysis about the subject area distribution) has been deleted. Besides, the **Section 3.2** (Analysis at the level of journal) has used time series to analyze all of the journal citation networks. Only the three main subject areas in IJF and JF have been analyzed in the **Section 3.2**, respectively. The analysis about the remaining top 10 subject areas has been provided in the **Appendix**. Please check it.

***Comment 17:***

*Same for journal citation networks -- just do the big 3, and put the rest in an appendix.*

***Our response to comment 17:***

Thank you for your advices. We have re-written the **Section 3.2.1** (Knowledge diffusion within the same subject area starts from IJF)and **Section 3.2.2** (Knowledge diffusion within the same subject area starts from JF). We use time series to replace splitting the data into 3 periods. Then both in IJF and JF, only the three main journal citation networks in the three main subject areas are provided. The remaining journal citation networks are offered in **Appendix**. Please check it.

***Comment 18:***

*Please don't use* *rotated tables -- they are too hard to read. It is almost always possible to redesign the table to avoid rotation.*

***Our response to comment 18:***

Thank you for your advices. All of the rotated tables have been modified and changed to non-rotated tables. Please check it.

***Additional response:***

According to your comments 16: time series plots rather than arbitrary splitting the data into 3 periods would be better. Therefore, in the **Section 3.2.1** (Knowledge diffusion within the same subject area starts from IJF) and **Section** **3.2.2** (Knowledge diffusion within the same subject area starts from JF), we delete all the content about the dynamic analysis through splitting the data into 3 periods.

In the previous manuscript, the **Section 3.2.3** (Knowledge diffusion of IJF outside the forecasting field) is conducted based on the dynamic analysis about the knowledge diffusion from IJF and JF in the **Section** **3.2.1** and **Section** **3.2.2**. In this revised manuscript, we delete all of the dynamic analysis in **Section** **3.2.1** and **Section** **3.2.2**. Therefore, the **Section 3.2.3** (Knowledge diffusion of IJF outside the forecasting field) is also deleted because it would be illogical to solely provide the content of the **Section 3.2.3** without the dynamic analysis about the knowledge diffusion from IJF and JF in **Section** **3.2.1** and **Section** **3.2.2**.

Thank you very much.

Yours Sincerely

Dejian YU

Shunshun Shi