Bigger Than You Thought:
China's Contribution to Scientific Publications and its impact on the Global Economy

I. Numbers of Articles
2. Citations and Quality of Articles
3. Effects on Economy: industry and innovation
4. Conclusions and More Research Needed

Richard B. Freeman, Harvard and NBER
Qingnan Xie, Nanjing University of Science and Technology
SBBI Economics of Science Seminar  Feb 1, 2019

“The one ring that rules them all … Knowledge
Basic research and frontier exploration.. must be done now,”
minister Wan Gang. Minister of Science and Technology, 2018
Abstract

This paper provides new estimates of China's contribution to global science that far exceed standard addressed-based measures of publications in international scientific journals.

Taking account of articles by Chinese researchers with non-China addresses as well as of researchers with Chinese addresses in journals in the Scopus database and of articles in Chinese language journals outside Scopus, adjusted for equivalence to Scopus articles, we attribute to China 36% of 2016 global scientific publications.

Increased citations to Chinese articles, particularly from non-Chinese addressed papers, and China's increased share of articles in top scientific journals shows that China advanced as much in quality as in quantity so that 37% of three year forward citations to papers published in 2013 were attributable to Chinese research.

China's new comparative advantage in knowledge was accompanied by increased share of output in knowledge intensive industries and innovation that impacts the future direction of research and of technological and economic development worldwide.
Our analysis begins with the largely English language Scopus database of journal articles, conference proceedings, books in natural sciences/engineering/math and social sciences.

Then we examine bibliometric data from the China National Knowledge Infrastructure (CNKI) that includes material published in China outside the Scopus data.

We analyze journal articles in “hard” natural/science/engineering/math and exclude social sciences.


Why journal articles? Better peer review and quality than conference proceedings and books.

Why only hard sciences? Because CNKI articles are more comparable to Scopus articles in those disciplines than in social sciences.
China's contribution to articles in Scopus

The standard measure of a country’s contribution to the scientific literature credits it for papers with its address, and for a fraction of papers with its address and those of other countries. Measured in this way S&E Indicators 2018 shows the shares of global articles, 2006-16
**Chinese articles in Scopus**

But this gives no credit to China for publications by Chinese researchers working at a non-Chinese address. Post-globalization, this diaspora research community is large: approximately 17 percent of *non-Chinese addressed* articles in 2016 had at least one Chinese-named author. To give some credit to China, we create a new measure of country contribution that weighs both addresses and names; Let $A$ be the number of authors with a given country address and $N$ the number of authors’ names associated with a country, we measure country $C$’s contribution to a paper as:

$$\alpha (Ac/A) + (1-\alpha) (Nc/N),$$

where $C$ subscripts denote address or national background/names and $\alpha$ is the weight given to addresses versus names. It varies from 1 (only addresses matter) to 0 (only names matter).
## Adjustment for Chinese Researchers at Non-China Address

### Table 1. Differences in Allocation of Credit for China

<table>
<thead>
<tr>
<th>Number of Chinese names with non-China address</th>
<th>Address based allocation of credit</th>
<th>Address and name based allocation of credit: $\frac{1}{2} (\text{China fraction of address}) + \frac{1}{2} (\text{China fraction of names})$</th>
<th>Difference, Equation (1) - address-based</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>0</td>
<td>$1/4 = \frac{1}{2} \left( 0 + 1/2 \right)$</td>
<td>3/12</td>
</tr>
<tr>
<td>2</td>
<td>1/6</td>
<td>$1/3 = \frac{1}{2}(1/6 + 1/2)$</td>
<td>2/12</td>
</tr>
<tr>
<td>1</td>
<td>2/6</td>
<td>$5/12 = \frac{1}{2} (2/6 + 1/2)$</td>
<td>1/12</td>
</tr>
<tr>
<td>0</td>
<td>3/6</td>
<td>$1/2 = \frac{1}{2} (1/2 + 1/2)$</td>
<td>0</td>
</tr>
</tbody>
</table>

Source: Authors’ calculations, as described in text.

Note: Example based on paper with six Authors, three with non-Chinese Names and Addresses and three with Chinese Names, by Number of Chinese Authors with non-Chinese Addresses.
Figure 1. Weighted Share of International Journal Articles to China, 2000–2016

- Total Chinese contribution = all Chinese-addressed + weighted share of international collaborations + \( \frac{1}{2} \) proportion of names on non-Chinese papers

- All China-addressed articles
- Proportion of Chinese names on non-Chinese papers
- Weighted share of international collaborations

Source: Scopus database.

Notes: Data classified by the year of publication, with papers weighted by proportion of Chinese address or names on the paper. Proportion of articles with non-Chinese addresses but at least one Chinese name estimated from a random sample of 20,000 Scopus articles with non-Chinese addresses in each year.
Figure 2. Proportion of Scopus Articles Associated with China

Source: Scopus database.

Note: Data calculated on basis of year of publication, with associated articles defined as having either a Chinese address or name.
These estimates of China's contribution to scientific articles misses CNKI journals not included in Scopus

If you think, as I first did, that the number of articles in these journals was small and must have fallen as Chinese researchers increased publications in Scopus journals, you are wrong!

In other countries own-language articles fell as English became the language of science. But not in China.

Between 2000 and 2016 China added 2.2 million papers with Chinese address by to international (Scopus) journals. Over the same period CNKI papers increased MORE rapidly than Scopus papers and fell in only a few fields and only in top Chinese institutions.

As largest 1st language in world, (and close to English as largest “1st + 2nd ” language), China could potentially challenge English as language of science.
Figure 3. Numbers of Science, Engineering, and Math Journal Articles in Scopus and CNKI, 1980–2016.

Source: Scopus and China National Knowledge Infrastructure (CNKI) databases.
Notes: Data calculated for journal articles only. The modest number of articles in journals covered in both databases are shown by the difference between the Total CNKI and CNKI-Overlaps lines.
Why the increase in CNKI publications?

- 1) Huge increase in R&D spending, expansion of faculty, PhDs whose careers depend on publications

- 2) Researchers in lower tier universities fill places in CNKI journals vacated by top researchers

- 3) China-specific content in CNKI: geology of China; China labor and business practices; patient experiences in China.

- 4) Publication required for master's and PhD graduates whose papers are one-off for degrees and rarely get cited.

- 5) Researchers “double-dip” publishing similar work in English and in Chinese
Fig. 4. Average Number of Chinese and English Language Articles in Three Tiers of Chinese Universities, 1990-2016

985 Universities: First Tiers

211 Universities: Second Tiers

Top 50 other Universities: Third Tiers
CNKI and Scopus classify research fields differently, so we took 12 large fields with same definition in CNKI and WOS. Seven CNKI fields keep growing (Oncology, Cardiology; Psychiatry; Pediatrics; Surgery; Geology, Environmental Sciences and Ecology). Five decline (Math, Microbiology, Optics, Metallurgy, Instrumentation).
Amalgamating Scopus and CNKI journal articles into a “Global” Whole

If the scientific content/impact of Chinese language papers was comparable to that of English language papers, the sum of Chinese articles in CNKI journals and of Chinese name and address weighted number of articles in Scopus, divided by the sum of all Scopus and all CNKI articles (minus articles in overlap journals) in 2016 would give us an estimate of China's contribution to global science in that year of 62 percent of scientific journal articles.

But CNKI and Scopus articles are not equivalent. Indicative of the quality difference, 44.6 percent of CNKI papers published in 2013 received no citations through 2016 compared to 29.0 percent of Scopus papers. Fewer scientists read Chinese than English, giving Chinese publications less scientific impact. Recognizing this, Chinese universities offer incentives for publishing in those journals (Arbritis and McCook, 2017; Quan et al., 2017), which induces many researchers to send their best work overseas, adding to the quality disparity.
<table>
<thead>
<tr>
<th></th>
<th>CNKI</th>
<th>Scopus</th>
</tr>
</thead>
<tbody>
<tr>
<td># of Articles in Nat Sci and Math 2013</td>
<td>1,640,565</td>
<td>1,550,883</td>
</tr>
<tr>
<td># of Articles with no citations (%)</td>
<td>750,560 (46%)</td>
<td>450,842 (29%)</td>
</tr>
<tr>
<td>Citations from own database</td>
<td>3,798,994</td>
<td>14,256,679</td>
</tr>
<tr>
<td>Average citations per paper</td>
<td>2.25</td>
<td>9.19</td>
</tr>
<tr>
<td>Estimated Citations from other database</td>
<td>19,987</td>
<td>3,798,994?</td>
</tr>
<tr>
<td>Total Citations from CNKI and Scopus</td>
<td>3,818,981</td>
<td>18,055,673</td>
</tr>
<tr>
<td>Average citations from CNKI and Scopus</td>
<td>2.33</td>
<td>11.64</td>
</tr>
<tr>
<td>Average pages of article</td>
<td>5.54</td>
<td>8.53</td>
</tr>
<tr>
<td>Refs given by articles</td>
<td>9.23</td>
<td>23.42</td>
</tr>
</tbody>
</table>
The Solution: Find an equivalence scale/exchange rate between CNKI and Scopus papers reflecting their relative importance.

Taking citations as the most accessible and widely used indicator of impact or quality, we transformed the number of missing Chinese papers into *Scopus equivalence papers* via a two-step procedure.

First, we calculated an exchange rate from the citations that Scopus and CNKI articles obtained in their own database. In 2013, a Scopus journal article averaged 9.2 citations from Scopus articles over the succeeding three years while a CNKI journal article averaged 2.3 forward citations from CNKI articles. This suggests a citation-based CNKI to Scopus exchange rate of approximately 0.25 (= 2.3/9.2).

But neither database includes citations received from publications in the other. Second, we correct for the omission of cross-database citations, by creating a REFERENCE data base and counting the cross-data base references to 2013 publications. Recognizing that a reference from X to Y is the forward citation that Y gets from X, we used the reference data to estimate the number of citations a 2013 Scopus article received from CNKI articles through 2016 and the citations a 2013 CNKI journal article received from Scopus through 2016.
The Result from this calculation is an estimated equivalence/exchange rate of about 20% -- 5 CNKI articles ~ 1 Scopus article

<table>
<thead>
<tr>
<th>Database</th>
<th>Number of total citations</th>
<th>Average citation per article</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Scopus</td>
<td>17,533,029</td>
<td>11.31</td>
</tr>
<tr>
<td>Scopus to Scopus</td>
<td>14,256,679</td>
<td>9.19</td>
</tr>
<tr>
<td>CNKI only Chinese articles to Scopus</td>
<td>3,276,350</td>
<td>—</td>
</tr>
<tr>
<td>CNKI non-Chinese articles to Scopus</td>
<td>3,276,350</td>
<td>—</td>
</tr>
<tr>
<td>Total CNKI</td>
<td>3,831,190</td>
<td>2.28</td>
</tr>
<tr>
<td>CNKI to CNKI</td>
<td>3,798,994</td>
<td>2.26</td>
</tr>
<tr>
<td>Scopus non-Chinese articles to CNKI only Chinese articles</td>
<td>32,196</td>
<td>—</td>
</tr>
</tbody>
</table>

Source: Scopus and China National Knowledge Infrastructure (CNKI) databases, tabulated by authors’ as described in the text.
Bottom Line: China's Share of Natural Sci-Engineering, Math Papers and Citations > twice its share of world population

Note: Based on estimated Scopus equivalence of CNKI articles as ratio of average citations of CNKI articles/average citations of Scopus articles = 0.20 in our calculations.
2. Citations and Quality of Articles
Cites to China address articles jumps from less than half world average citations to average: 1996–2012 (SEI, 2016)
Share of Cites in top 1% of cited papers shows similar pattern

Note: SEI2016 citation data are based on all citations made to articles in their publication year and all following years and are normalized by subfield and publication year. WOS data are this high cited paper received enough citations to place it in the top 1% of its academic field based on a highly cited threshold for this field and publication year, as of November/December 2016.
Source: National Science Board; WEB OF SCIENCE.
Our Calculations: World Citations and Share of China-Linked Scopus Papers in World Citations, 2000 and 2013

Average citation/World

- Articles with only Chinese addresses: 2000: 0.29, 2013: 0.70
- Articles with both Chinese and other countries’ addresses: 2000: 1.05, 2013: 1.78
- Articles with no Chinese address but Chinese names: 2000: 1.88, 2013: 1.45
- China’s weighted average citation: 2000: 1.00, 2013: 0.90

Weighted share of total citation

- Articles with only Chinese addresses: 2000: 1.18, 2013: 11.07
- Articles with both Chinese and other countries’ addresses: 2000: 0.22, 2013: 1.64
- Articles with no Chinese address but Chinese names: 2000: 5.99, 2013: 6.76
But Citations Have a Homophily Problem as a metric of quality

People cite papers of people like themselves more than of others:
- Women disproportionately cite papers by women.
- Swedes disproportionately cite papers by Swedes/other Nordics.
- Harvard people disproportionately cite papers by Harvard people.
- Boston people disproportionately cite papers by Boston people.

If, as is the case, Chinese researchers are more likely to cite Chinese papers than other researchers, then the homophily preference will raise Chinese citations as the Chinese share of papers increases.
How much of Greater Citations is Homophily of Citing x growth of China papers vs Improved Quality of papers?

When China contributed few papers, citation homophily --> undervalue China research. Huge increase in # China papers → more citations. To get better fix on citations as quality, examine 3 yr non-China citations to China papers.

Ln increases
CC-- +0.63
NC-- +0.46
so ~60% quality

Cites to -China papers by year of publication

China to China Cites Per Paper

Non-China to China Cites Per Paper
Appendix C: Fraction Weighted Share of Papers in Nature and Science, for Chinese Addresses and Names on Articles, 2000 and 2016

China in *Nature*

- 2000: 7.14%
- 2016: 17.15%

China in *Science*

- 2000: 7.91%
- 2016: 16.89%

Legend:
- Purple: Weighted articles with non-Chinese address but Chinese names_Nature
- Green: Weighted articles with non-Chinese address but Chinese names_Science
- Red: Weighted articles with Chinese address_Nature
- Blue: Weighted articles with Chinese address_Science
3. Effects on Economy in Knowledge Economy

RD/Knowledge Model: Science $\rightarrow$ Greater Stock of Useful Knowledge $\rightarrow$ Tech Development/change $\rightarrow$ Innovation $\rightarrow$ Higher Productivity; improved attributes of products

- Industry mix should shift to knowledge intensive production
- Share of global production and exports in knowledge intensive industries should shift.
- Migration --End of *brain drain* that strengthened advanced country knowledge advantage. China attracts more return migrant S&E workers and knowledge-intensive FDI, vide Apple and Google 2017 plan to open research facilities in China.
- Mode of economic development – end of *North-South model* where advanced country monopoly of R&D- technology and innovations boosts wages/productivity vs developing country.
Measuring these expected effects

- Standard economic measures of output and trade in knowledge-intensive sectors of economy.
- Increased domestic patents but quality of Chinese SIPO patents not comparable to USPTO, JPO, EUPTO and many new products/processes are not patented.
- Increased overseas patents – USPTO
- Best measure is “innovations” but lack national statistics on new products/processes in regular GDP or satellite knowledge/ intangibles what we have are indicators, many of which are per-capita basis.
- Cross-country comparisons difficult because business folk in different countries appear to report innovations differently.
<table>
<thead>
<tr>
<th>Year</th>
<th>China’s share of world production</th>
<th>China’s share of world exports</th>
<th>Share of Chinese production</th>
</tr>
</thead>
<tbody>
<tr>
<td>High-Tech industries</td>
<td>0.04 0.17</td>
<td>—</td>
<td>0.27 0.35</td>
</tr>
<tr>
<td>High-tech manufacturing (aerospace, communications and semiconductors, computers and office machinery, pharmaceuticals, and scientific instruments and measuring equipment)</td>
<td>0.06 0.24</td>
<td>0.1 0.24</td>
<td>0.03 0.03</td>
</tr>
<tr>
<td>Information communication technology (communications, computers, and semiconductors)</td>
<td>0.06 0.28</td>
<td>0.12 0.36</td>
<td>0.02 0.02</td>
</tr>
<tr>
<td>Medium high-tech manufacturing (motor vehicles, electrical machinery and apparatus, chemicals excluding pharmaceuticals, railroad and other transportation equipment, and machinery and electrical equipment)</td>
<td>0.19 0.32</td>
<td>0.07 0.2</td>
<td>0.08 0.09</td>
</tr>
</tbody>
</table>
Patent Growth “faster than a speeding bullet”

Number of patent application 1990-2015
restricted to domestic applicants filing to domestic office

- China
- US
- Korea
- Japan

Data source: WIPO Statistics Data Center
In May 2014 US vice president Joe Biden dismissed China's ability to turn its S&E expertise into economic innovation: “I challenge you, name me one innovative project, one innovative change, one innovative product that has come out of China”.

To answer Biden's challenge, here are four “top ten innovations” at Las Vegas' World's Consumer Electronics Fair 2018.

- an underwater drone (Beijing);
- a light electrical bicycle (Shenzhen);
- a fingerprint sensor for smart phones (Dongguan);
- a virtual reality headset (Lenovo).

China's share of global innovations undoubtedly varies across sectors and technologies but nothing “wierd” in CEF2018. Test in its new abilities will be in meeting goal for major role in artificial intelligence: already producing most papers in AI.
“Soft evidence”


KPMG's 2018 global technology innovation survey of technology industry leaders puts China 2nd to the US as the most promising market for tech breakthroughs.

Other evidence says still a way to go.

2018 Global Innovation Index ranks China 17th, up from 39th in 2013 but it uses per capita indicators with the 2018 leaders being Switzerland, Netherlands, and Sweden.

Forbe's list of 100 most innovative companies has seven Chinese firms; BCG lists three Chinese firms in its 50 most innovative companies.
4. Conclusion

- China has turned its investment in science and engineering into comparative advantage in knowledge creation that contributes to economic innovation.

- To the extent that knowledge is the “one ring that rules them all”, the way China deploys its scientific resources will be a key driver of scientific, technological, and economic progress in the foreseeable future.

- We need more evidence on where allocating sci-tech resources can best help China and world advance in science and economy.

- To deal with challenges that face us, we need more knowledge on how economics and policies shape science and how science-based innovations shape the economy.

To paraphrase Horace Greeley’s advice to Americans as the US expanded to California “Go West, young man, and grow up with the country”3 science is going East and will grow up with China. So too should a substantial share of our research.