Standing on the Shoulders of Chinese Giants? A Study on International Biases in the Process of Global Scientific Knowledge Diffusion

by

Valerie Y. Kuo

B.S. Mechanical Engineering
Massachusetts Institute of Technology, 2006

SUBMITTED TO THE SLOAN SCHOOL OF MANAGEMENT IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE
AT THE
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

FEBRUARY 2012

Signature of Author: ______________________
MIT Sloan School of Management
December 15, 2011

Certified by: ______________________
Fiona E. Murray
David Sarnoff Professor of Management of Technology
Thesis Supervisor

Accepted by: ______________________
Ezra Zuckerman Sivan
Nanyang Technological University Professor
Chair, MIT Sloan PhD Program

Standing on the Shoulders of Chinese Giants?
A Study on International Biases in the Process of Global Scientific Knowledge Diffusion

by

Valerie Y. Kuo

Submitted to the Department of Management on December 15, 2011 in Partial Fulfillment of the Requirements for the Degree of Master of Science in Management

ABSTRACT

Recent research has shown rapid growth in the number of Chinese-authored publications over the last decade. However, while the volume of publications has increased dramatically, the total number of citations – a popular measure of quality or impact – to such articles has risen at a decidedly slower pace.

Three possibilities might explain the cause of this slow growth in citations. First, it could be that most Chinese-authored papers fall in the left-hand tail of the quality distribution. This is likely at least a partial explanation for the lag in quality. However, we believe that other factors also contribute to the quality gap. A second explanation for the citation lag is that Chinese scientists, while prolific, are simply not publishing in the most popular fields or in the most cited journals. If true, this scenario would exemplify a long-argued shortcoming of citation metrics and provide evidence supporting the policies and progress of Chinese science. A third possibility is that there exists a bias against Chinese science and Chinese-authored papers are undervalued and underutilized.

To explore these possibilities, we collected citation and other key data on scientific publications from the highly prestigious and rigorously screened Nature and Nature-derivative (which we will refer to as Nature X) journals. In addition to conditioning out publication quality, we employ coarsened exact matching (CEM) to match Chinese-authored articles with similar US-authored articles before estimating the effect of Chinese-authorship on citations received.

Our results suggest that a bias does exist against Chinese-authorship. While this paper identifies the existence of a citation gap, it does not attempt to identify the source of this citation gap. Thus, future research could focus on uncovering the exact mechanisms by which this phenomenon occurs. Doing so will no doubt inform policy and institutional practices to the benefit of scientific progress and, ultimately, societal welfare.

Thesis Supervisor: Fiona E. Murray
Title: Professor of Management

Acknowledgements

I would like to thank Professors Fiona Murray and Yasheng Huang for their guidance and support throughout this academic research process.

I would also like to extend my gratitude to the general Technological Innovation, Entrepreneurship, and Strategic Management (TIES) group at Sloan for providing a challenging and stimulating experience over the past two years. As difficult as it has been at times, I have never learned more or been as intellectually challenged as I have in this program. Thank you to all the faculty members, administrators, and fellow students who have contributed to this unique experience.

Finally, I am forever grateful to my family and friends for their continued and unwavering support.

Standing on the Shoulders of Chinese Giants?

I. Introduction

> *"Everybody is looking at China and saying, if we don't lift our game, China is going to eat our lunch economically because the amount they are investing in science, technology and innovation, while it has not yet reached anything like our level, is rising very quickly"* (Connor, 2011) .
>
> - *John Holdren, Director of the White House Office of Science and Technology Policy*

Over the last decade China has been increasingly recognized as a major contributor to global scientific innovation. One of the most popular methods used to measure this change is through the analysis of data on scientific publications originating in China (this is referred to as bibliometric data analysis). Recent research has shown rapid growth in the number of Chinese-authored publications over the last decade. However, while the volume of publications has increased dramatically, the total number of citations – a popular measure of quality or impact – to such articles has risen at a decidedly slower pace. This paper is an attempt to understand the forces contributing to this phenomenon of unequal growth.

Three possibilities might explain the cause of this slow growth in citations. First, it could be that most Chinese-authored papers fall in the left-hand tail of the quality distribution. This is likely at least a partial explanation for the lag in quality. Relative to Western countries, China has only recently begun to build and revitalize its science and technology infrastructure, and a focus on high impact science has not yet been indoctrinated throughout the country. Whether or not the current results were intended by the Chinese government, understanding the outcome of its multi-pronged approach will be informative to both the Chinese government, as well as governments of other developing countries. Articles from both the popular press and from academic literature support this explanation. However, we believe that other factors also contribute to the quality gap.

A second explanation for the citation lag is that Chinese scientists, while prolific, are simply not publishing in the most popular fields or in the most cited journals. The practice of citing existing publications often differs across fields of study, and Chinese scientists may – by chance – be focused in these areas receiving low average citation counts. If true, this scenario would exemplify a long-argued shortcoming of citation metrics and provide evidence supporting the policies and progress of Chinese science. In addition, it may reduce any bias against Chinese science, a view that is often perpetuated by the popular press.

A third possibility is that there exists a bias against Chinese science and Chinese-authored papers are undervalued and underutilized. A prevailing theory in economics cites knowledge diffusion as a key component to economic growth. Understanding the mechanisms behind the diffusion of scientific research can inform policy or research practices to increase the efficiency of this knowledge transmission. If this is the case, we will have brought light to yet another source of inefficiency in the process of knowledge diffusion, but one that education and awareness can likely mitigate. To explore this possibility, we collected citation and other key data on scientific publications from Nature and Nature-derivative (which we will refer to as Nature X) journals. These journals are highly prestigious and are notorious for their rigorous submission and screening processes. As a result, the quality of Nature- and Nature X- published papers can be viewed as roughly equal. In addition to conditioning out publication quality, we employ coarse exact matching (CEM) to further match Chinese-authored articles with similar US-authored articles and then run several negative binomial regressions to estimate the effect of Chinese-authorship on the number of article citations received by a paper.

Our results suggest that a bias does exist against Chinese-authorship. This is somewhat unexpected given the prestige awarded Nature and Nature X publications, particularly since science as an institution champions the practice of objectivity. At the same time, however, social, institutional, and geographical factors have been found to play a role in knowledge diffusion and, specifically, scientific citation practices. While this paper identifies the existence of a citation gap, it does not attempt to identify the source of this citation gap. Thus, future research could focus on uncovering the exact mechanisms by which this phenomenon occurs. Doing so will no doubt inform policy and institutional practices to the benefit of scientific progress and, ultimately, societal welfare.

Section II of this paper provides an overview of existing literature and outlines our hypotheses. Section III discusses the nature of the data and our methodology. Section IV gives results, and Section V ends with our conclusions and thoughts for future research.

II. Background

Chinese Policy in Science & Technology

Much of the S&T infrastructure in China today was implemented after the Cultural Revolution (1966-1976), which – for all intents and purposes – destroyed the systems and institutes that were in place prior to the ascendancy of Mao Zedong and the Communist Party of China. Since Deng Xiaoping officially launched the "Four Modernizations" in 1978[1], the government has hatched and implemented dozens of policies and poured its resources into building the foundation of what is, today, an acknowledged scientific force. Besides increasing overall funding for science and technology, the Chinese government has put heavy emphasis on the rebuilding of its science academies and universities, which today are China's main producers of scientific publications published in English-language academic journals. The remainder of this subsection summarizes key policies and the subsequent development of the Chinese Academy of Sciences – the premier scientific academy within China – and the universities within the higher education system.[2]

The Chinese Academy of Science (CAS) was established on November 1, 1949, and its mission and structure was developed under strong Soviet influence. At the beginning, there was a strong presence of first-rate Chinese scientists, often trained abroad, who returned home and made serious contributions to modern Chinese science. However, the Cultural Revolution destroyed much of what was previously built, and policies enacted in later decades were necessary to rebuild and re-invigorate Chinese science. In particular, the 1980s brought several reforms which established "competitive, project based national programs for research and institutional improvement (Suttmeier, Cao, & Simon, 2006, p. 81)," such as the National Natural Science Foundation of China, 863 Program, National Key Laboratory Program, and National Engineering Research Center Program. (Suttmeier, Cao, & Simon, 2006; Chinese Academy of Sciences; Liu & Zhi, 2010)

Further reform under the Knowledge Innovation Program (KIP) had a particularly great impact on the CAS. The KIP, which was first implemented in 1998, whittled down 120 inefficient

[1] The Four Modernizations were major reforms in the areas of agriculture, industry, national defense, and science & technology. They were intended to mold China into a competitive force with the rest of the modern world.
[2] See Fensterheim (2009) for a thorough summary of the history of the Chinese Academy of Sciences and the Higher Education Institutes in China.

institutes (which had overlapping projects, overabundance of administrative personnel, dead-end research, low employee productivity, etc.) into 90 re-organized, well-run institutes by 2006. The KIP also changed funding structures so that the institutes would have greater autonomy in the management of their own research. In parallel, the KIP has been supplemented by a peer evaluation system which created dialogue between leading Chinese and foreign scientists. (Suttmeier, Cao, & Simon, 2006; Liu & Zhi, 2010)

The KIP also enacted the "100 Talent Program," which sought to recruit high potential scientists, many from abroad, by offering enticing incentives such as higher salaries, better benefits, and brand new equipment and laboratories to run (Suttmeier, Cao, & Simon, 2006). These new appointments no longer promised lifetime tenure but instead began to implement evaluations early in the scientist's career. Salary structures were modified to account for the added administrative responsibilities of leading scientists, and merit-based compensation was supposedly implemented for high performance. The CAS simultaneously began to shift focus from publication quantity to quality: "Under this system, the funding of the next year is highly related to achievements of the current year. Evaluators are also invited from outside CAS in order to eliminate bias. Evaluations are carried out at all levels, from CAS headquarters down to each individual's performance" (Liu & Zhi, 2010, p. 334).

However, not all of these changes worked as efficiently as expected. In particular, Suttmeier et. al. (2006) makes the following observation:

> *"Programs to improve the talent pool by recruiting Chinese scientists working abroad to return to China also have not escaped some of the problems of fraud and corruption which have plagued Chinese science recently. In some cases, the high salaries and attractive material incentives used in these programs have been abused. Researchers have enjoyed the salaries without taking their research responsibilities seriously, that is, without fulfilling the obligations of appointments, while their employing institutions have been satisfied to use the names and publications by these 'star scientists' to improve their evaluations and thus qualify for increased funding."* (Suttmeier, Cao, & Simon, 2006, p. 88)

Currently, the CAS employs over 50,000 people across its 12 branch offices, 103 institutes, over 100 key laboratories and engineering research centers, and 1,000 field stations (Chinese Academy of Sciences). The CAS, along with China's National Natural Science Foundation, Chinese Academy of Social Sciences, and Chinese Academy of Engineering, are under the jurisdiction of the Ministry of Science & Technology (MoST). Both the MoST and Ministry of Education (MoE) – in charge of the nation's universities – report to the National Steering Group on Science, Technology and Education, a committee of the State Council (Fensterheim, 2009).

Like the CAS, China's higher educational institutes (HEIs) were in need of reform and restructuring post-Cultural Revolution. In 1985, the central government engaged in decentralization, passing more responsibilities to the local government and effectively giving HEIs more managerial autonomy. However, the central government required all universities to house a handful of activities: teaching, research, business, and social services. In 1993, further reforms developed the "user-pays" system, the same system employed largely in the West. Under this system, students pay for tuition, and this reform led to the mass marketization of higher education in China during the 1990s (Bai, 2006). Project 985 – launched in 1998 – increased the amount of funding for conducting research, improving facilities, and engaging in international collaboration at top universities. In 1995, China launched its "211 project" which focused on developing its best 100 universities and turning them into world-class institutes. These top universities collectively utilize approximately 70% of the State's scientific research funding and house almost all of the State's key laboratories (Over 10 billion yuan to be invested in "211 Project", 2008).

As the implementation of the 211 project suggests, China began to focus more on quality over quantity in the 1990s. According to Li et. al (2008):

> *"It is now accepted as important for universities and related institutions to achieve publication in journals of good ranking and what is generated by publication citations counts equally for Chinese scholars in appointment, maintenance of position, and promotion. Indicators of educational attainments in terms of international rankings across countries, publications of papers, and citations feed directly into annual performance indicators for Chinese faculty in*

an ongoing process which goes substantially beyond the once in a lifetime tenure system outside China." (Li, Whalley, Shunming, & Xiliang, 2008, p. 12)

However, this emphasis on publications and citations has not necessarily improved university publication quality as much as one might hope. One unintended consequence of this new system is that scientist may try only to achieve the bare minimum of the requirements (Jacobs, 2010; Yimin, 2001). The per-article incentive system at universities has effectively defined a minimum measure of "quality" by which scientists must achieve to maintain their post:

> *"The campaign for more international publications, especially in journals included in the Science Citation Index, a bibliometric database compiled by Thomson Reuters, has an unintended consequence — institutions of learning have placed more emphasis on quantity, and assessed, promoted and rewarded their scientists accordingly. When a scientist has difficulty fulfilling the required quantity for the position legitimately, he or she is likely to divide the research into 'the least publishable unit,' or even take a detour.* (Cao, 2010)

This culture of "publish or perish" has also anecdotally increased levels of academic fraud (Xin, 2006; Dickson & Hepeng, 2006; Cao, Climate for scientific misconduct in China, 2010; Jacobs, 2010). One particularly publicized act of academic misconduct occurred in December of 2009, when the editors of *Acta Crystallographica Section E* retracted 70 articles by Chinese authors due to suspected fraud (Harrison, Simpson, & Weil, 2010). These acts of misconduct, which have been documented as early as the 1990s (Li & Xiong, 1996; Wang, 1999), may well have caused a bias to develop against Chinese-authored publications.

Citation Number vs. Citation Impact

A recent report published by the Royal Society (see Figure 1) ranked China second – behind the US – in the number of English-language scientific publications produced annually. Furthermore, it projected that China will outpace the United States in volume by as early as 2013. However, the report also found that China lags significantly behind its western counterparts when the measure is one of impact – the number of citations to Chinese-authored papers – rather than sheer quantity (Royal Society, 2011; Shukman, 2011; Jha, 2011).

But a simple count of citations does not incorporate subject-level differences in citation patterns, meaning that China could simply be publishing in subjects that are less likely to get cited, but that are no less important or impactful. King (2004) takes this possibility into account by calculating the average citations per paper from a country and then normalizing the average across subjects. The results in Figure 2 show that China ranked near the bottom of the scale in 2002 and further corroborates claims that Chinese science is not progressing as quickly as originally believed. A third popular method used to analyze scientific performance or impact is to look at the composition of countries among the most highly cited papers. King (2004) isolates the top one percent of the most highly cited publications published between 1997 and 2001 and finds that, while Chinese publications made up roughly 3.18% of the world share in publications and 1.56% of citations, they only contributed 0.99% to the pool of elite (top 1%) publications (Figure 3).

In summary, the number of Chinese-authored articles published each year has grown exponentially, but measures of impact suggest that the majority of these publications fall in the left-tail of the quality distribution. One possible explanation for this is the "publish or perish" culture, which forces academics to focus on the quantity rather than the quality of their work. Even worse, this pressure has led to incidences of academic fraud, which may have led the international community to develop an undeserved bias against high-quality, Chinese-authored publications. Whether or not there is an actual bias against such publications, however, has not been studied to any great extent. Thus, the purpose of this study is to identify whether there is, indeed, room for improvement in this realm of international scientific exchange. The next logical question one might ask why scientific exchange should be a topic of interest. The next section will discuss this in detail.

Knowledge Diffusion

According to the widely accepted economic theory of endogenous growth, technological progress is the engine of economic growth, and the total stock of knowledge is in turn the key component of this progress. The amount of human capital devoted to scientific research is one factor contributing to this stock (Romer, 1990). Another factor, argued to be even more important (or at least more appealing) because it is easier to manipulate through policy, is the rate or effectiveness of knowledge diffusion (Aghion & Howitt, 1992).

One important mechanism by which scientific knowledge diffuses is through publications which are read widely by the academic community – a community which has its own set of norms that incent scientists to publish or share their best work (Merton, Priorities in Scientific Discovery: A Chapter in the Sociology of Science, 1957; Dasgupta & David, 1994)[3]. Basically, academic scientists have developed a first-to-reveal or priority-based system which rewards them with status and reputation for publishing and sharing their ideas/findings. Thus, only by forgoing control of and quickly sharing their findings are they able to gain status and reputation, which they innately value more than monetary gains. In addition to publishing, the norm of openness – or communism – in science also compels scientists to cite the work of their peers which have influenced them in their own research (Merton 1957; Dasgupta & David 1994; Sorenson & Fleming 2004)[4].

While academic journals are an effective mechanism through which high-quality research can be disseminated, less discussion has been devoted to understanding the citation patterns, or the use, of such easily accessible information. In other words, knowledge diffusion cannot occur unless the access to information is accompanied by the use of this information. The sociology and innovation literature identifies some theories which may explain patterns (and inefficiencies) in use of prior knowledge. One stream has studied – largely through the analysis of patent citations – the geographical localization of knowledge. The findings from these studies suggest that the effect of knowledge diffusion is often strongest in geographically close locations (Jaffe, Trajtenberg, & Henderson, 1993; Audretsch & Feldman, 1996). This is likely due to the fact that information still travels most often from person-to-person, and the nature of relationships is that they tend to be localized. Less research has been conducted to study the bounds of this geographic localization. In this study, we expect to see evidence of a preference for own-country publications.

A second explanation for the citation differential is summed up by the literature on what is known as the Matthew effect (1968), named after a line from the Gospel of Matthew:

[3] Other mechanisms of scientific knowledge diffusion include training of students, peer-to-peer interaction, conferences, etc. (Sorenson & Fleming, 2004)

[4] See section 2 of Sorenson & Fleming (2004) for a more detailed discussion of the norms in science which shape the desire to publish.

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, New Revised Standard Version

In the context of academic publications, the Matthew effect predicts that famous scientists will receive rewards in the form of recognition and resources, which enables them – above others – to accomplish more and receive even greater rewards. In the particular institution of science, one way in which this recognition manifests is through the over-allocation of citations to a scientist's publications. However, scant research has been conducted on the Matthew effect, particularly in the context of countries. The literature that does exist indeed indicates the presence of the recognition misallocation (Bonitz & Scharnhorst, 1999; Bonitz, Ten years Matthew effect for countries, 2005; Bonitz, Ranking of nations and heightened competition in Matthew Core journals, 2002). In light of the academic scandals in China that have been publicized over the last decade, in addition to the fact that China only recently (relative to much of Europe and the US) began to revitalize its science and technology platform, it is likely that a cognitive bias has developed in the global scientific community – a bias which has been amplified despite the obvious improvements in China's research infrastructure and policies.

III. Empirical Strategy

Citation Data as Measure of Impact

The use of citation data to measure various aspects of science is not new. Some of the earliest studies measured impact by a simple count of publications published annually by an individual, department, or country. However, most publications are never read or are very low in impact (left-skewed in quality distribution), so grouping those publications with the higher impact publications in a simple count would be misleading. Instead, researchers began looking at citation patterns to publications. Since a citation to a paper indicates that it was read and used by the scientist, it is a more accurate measure of impact. A simple count of citations to a paper can measure that paper's impact, or similarly the average citations per paper for a group, department, or country. Another popular measure is a group, department, or country's share of total citations. In other cases, such as in this paper, the interest is in comparing the performance of elite groups from two different countries. We isolate an elite group of publications (*Nature* and *Nature* derivatives) and compare the differential in the number of citations received. It is also possible to go one level deeper and examine the impact of the *citing* papers. For instance, we will look at the Journal Impact Factor (JIF) of the citing papers, where JIF represents the moving 2-year average of citations to a paper within a particular journal. (Elkana, Merton, & Zuckerman, 1978; Stigler, 1979; Van Raan, 1988; Adams & Griliches, 1996; Thomson Reuters, 2008)

Since Jaffe, Trajtenberg & Henderson (1993) published their seminal piece on the geographic localization of spillovers, the use of citation data has exploded in the diffusion literature. The popularity of citation metrics has grown due to its many attractive qualities. Despite Krugman's (1991) claim that knowledge flows are invisible and therefore cannot be studied, Jaffe et. al. (1993) proved that they could, indeed, be collected, traced, and measured by following the "paper trail" of citations. Citation data is both well-documented and easy to access. Additionally, the institution of science has forces which make citation data reliable indicators of knowledge flows and impact. Not only are scientists incentivized to publish as quickly and as openly as possible in order to gain recognition, but a culture exists where a scientist cites prior work that has influenced him in the pursuit of his own research. Thus, a reasonable assumption can be made regarding the ability of citation data to represent the current state and use of scientific knowledge. (Merton, 1957; Merton 1942; Dasgupta & David, 1994)

Naturally, however, there are also weaknesses to citation metrics. Citations metrics, while easy to use, do not capture all the activities which contribute to knowledge spillovers. Activities such as attending conferences, talking to one's peers, patenting, or mentoring and teaching are other pathways through which knowledge can flow and which are not capture through publication-to-publication citations. It is also possible that a scientist might simply forget to cite a work of significant contribution, or perhaps the work is so ubiquitous (such as Einstein's theory of relativity) that it has become obvious and needs no citing. Third, some scientists may cite their friends or co-workers just to be nice or to do somebody a favor when it is unmerited. In addition, the time between a publication and any impact it has may well vary. Some papers might have immediate impact, while others may take years or even decades to bear fruit. (Moed, Burger, Frankfort, & Van Raan, 1985; Vinkler, 1988; Jaffe, Trajtenberg, & Henderson, 1993; Moed, De Bruin, & Van Leeuwen, New bibliometric tools for the assessment of national research performance: Database description, overview of indicators and first applications, 1995)

Thus, when working with citation data, it is important to keep in mind these weaknesses and to avoid drawing overly ambitious conclusions from any results. But this can be said of any form of analysis, and no methodology will be perfect. With that in mind, we will move on to a description of the data and, subsequently, the methodology used in this study.

Data

The data used in this study was collected through Thomson Reuter's Web of Knowledge research platform and in two stages. In 2008-2009, data was gathered through a web scraping script[5] on over 130,000 journal articles published between 1979 and 2007 in Nature and Nature auxiliary journals.[6,7] Each root article[8] observation contains identifying data such as the journal title, article title, authors, author addresses, article key words, date of publication, etc.[9,10] Since we are interested in the performance of Chinese-published articles with respect to a Western benchmark, we retained only those articles that have a Chinese or a US address listed under the first-author field. Additionally, since we are interested in the impact of modern or current Chinese science, we dropped all articles published prior to the year 2000 (many key programs were initiated in the 1980s and 1990s, so starting with the year 2000 seems to provide an appropriate amount of lag time). Upon completion of these deletions, the database contained roughly 11,000 root articles, with only 135 Chinese, first-authored articles. Given the highly unbalanced nature of the data, each Chinese article was then matched with US articles appearing in the same issue, and unmatched US articles were dropped from the database, leaving 1,050 US articles.

While this initial database had the collective number of citations to each root article (our dependent variable of interest), it did not contain information on the year or time of each citation that the article received – a key piece of data given that the incidence of citations tends to occur as a function of the article age or vintage (Hall, Jaffe, & Trajtenberg, 2001). In order to control

---

[5] This script was written by Devin Fensterheim, who at the time was a Masters Student at Sloan.
[6] Does not include book reviews, editorial material, letters, proceedings papers, etc. (document type limited to articles), nor does it include books, books in series, etc. (publication type limited to journal)
[7] A list of all Nature auxiliary journals can be found in the Appendix
[8] A root article is the originating point of interest. An article that cites a root article (or publication) is referred to as the citing article (publication).
[9] See *Web of Science 8.0* (The Thomson Reuters Corporation, 2009) for a detailed explanation of the Thomson Reuters data collection methods, its citation databases, and the web search procedure.
[10] See Appendix for the full list of variables

for the age at citation, additional data was collected by hand through the Thomson Reuters Cited Reference Search. Identifying information for each article on the order of title, year of publication, and authors, was entered into the web search tool for each root article (aka the article of interest), and a list of publications which cite this root article is returned. The data for each citing publication contains the same variables as originally obtained for each root article (e.g. journal title, article title, authors, author addresses, etc.). The total sum of citing publications is 169,710, which includes all document and publication types. The primary citing publication of interest is the journal article, and within this citing data, there are 116,874, or approximately 69% of the sample. By collecting this data, it is possible to subtract the year of publication of the root article from the year of publication of the citing publication to obtain the age of the root article at citation. In addition, the level of detail in the citing data allows for more detailed analysis regarding who and what are most impacted by Chinese and US-authored Nature publication.

Finally, a unique list of publications from the citing journals was culled and categorized as either a high JIF or a low JIF citation. Publications were labeled as high JIF if they were listed as having a top 500 JIF score as defined in Thomson Reuters' ISI Web of Science database.[11] The Thomson Reuters database only has JIF scores available for the years 2001-2010, so analysis using this JIF data will be constrained to a smaller subset of citing publications (76,755 total). Since the JIF score is a moving average and changes every year, we were careful to collect the JIF score for each journal in every year.

Additional post-processing was also performed in Python and Stata to parse out key information relating to the country of origin for each observation. Finally, the data was expanded to a long format such that each observation represented an article-year. For instance, an article X published in 2005 would transform into X-2005, X-2006,...X-2011. This format is optimal for our regression analysis (documented in detail later), which uses as the dependent variable the number of citations received each year (and controlling for age and year of publication). In other words, our data can be described as a time-series, cross-section panel, where observations take the form of an (root) article-year. See Tables 1 – 3 for variable definitions and descriptive statistics on the entire sample.

[11] Ideally, the JIF score of each citing publication would have been searched and recorded; however, the high number of unique publications, multiplied by a ten-year timeline, would render this task too time consuming. Additionally, not all journals are assigned a JIF score.

Does Chinese Science Have an Impact?

In a perfect world, papers would be submitted anonymously and the country of origin for first authors would be randomly assigned to China and the US. Data would be collected at year x for all Nature articles published in year x. We would then run a difference means analysis by calculating the average number of citations received to each paper at year x + t (where t is something like 5 years or 10 years) separately for Chinese-authored and US-authored papers and comparing the difference between the two. This difference would reflect the bias against (or for) Chinese-authored papers. However, the world of observational data is decidedly imperfect. Since there is no random assignment, we have to account for the fact that Chinese-authored papers might differ systematically from US-authored papers on certain characteristics. For example, Chinese scientists might have a tendency to collaborate less frequently or have fewer co-authors than US scientists. Since teams tend to produce higher impact papers (Wutchy, Jones, & Uzzi, 2007), any additional citation received by US authors may be attributed to the systematic difference in team size rather than the country of origin.

One way to deal with this and similar other issues is to employ a matching method in order to "match" a treated (Chinese-authored) paper with a control (US-authored paper). The purpose of matching is elegantly explained in Iacus, King, & Porro (2011):

> *Matching is a nonparametric method of controlling for the confounding influence of pretreatment control variables in observational data. The key goal of matching is to prune observations from data so that the remaining data have better* balance *between the treated and control groups, meaning that the empirical distributions of the covariates (**X**) in the groups are more similar. Exactly balanced data mean that controlling further for **X** is unnecessary (since it is unrelated to the treatment variable), and so a simple difference in means on the matched data can estimate the causal effect; approximately balanced data require controlling for **X** with a model (such as the same model that would have been used without matching), but the only inferences necessary are those relatively close to the data, leading to less model dependence and reduced statistical bias than without matching.* (pg. 1)

One particular method – coarsened exact matching (CEM) – has been brought to light only very recently (Blackwell, Iacus, King, & Porro, 2009) and is one solution to the "curse of dimensionality" issues most often associated with exact matching procedures. To our knowledge, it has been applied only once thus far in the field of knowledge diffusion (Azoulay, Zivin, & Sampat, 2010). In their paper, Azoulay et. al. employ CEM to match articles authored by superstar scientists who were relocated to a new laboratory (treatment) with "stayer" superstars (control, the scientists who were not relocated) to study the effects of relocation on knowledge diffusion patterns such as co-authorship behavior. It is this method which we employ in the current study. The steps to CEM are simple: (1) "coarsen" the criteria for the covariates (variables which you want to match with the controls); (2) perform exact matching on the coarsened covariate criteria; and (3) drop the controls from the sample which did not find a match (King G. , 2010; Blackwell, Iacus, King, & Porro, 2009). This study uses the CEM function in Stata, developed by and detailed in Blackwell, Iacus, King, & Porro (2009).[12]

We identify controls based on the following set of time-invariant covariates: (1) exact match on the issue, which includes the specific journal, the publication year, and oftentimes more granular levels of date, such as quarter or month (e.g. Jan 27 issue of Nature in year 2000) (2) number of authors (coarsened into groups of 1, 2, 3- 4, 5-10, or >10 authors), and (3) number of countries (coarsened into groups of 1, 2, >3 countries). Using these criteria to conduct the CEM method, 589 of the 1020 (58%) Chinese observations (Chinese-authored article-years) were matched to 1235 of the 8330 (15%) US observations. Since the matching procedure was not exact (i.e. coarse), there still remains some covariate imbalance between the control and treatment groups in addition to a time-varying component to the data. The next logical step, then, is to develop a statistical model to control for these various elements. See Table 4 for descriptive statistics on these matched groups.

One important feature of the data in this study is that it falls under the category of count data, meaning that it takes on integer values equal to or greater than zero. Traditional OLS regression on count data, as is the case with citations, often results in estimates that are biased, inefficient, and inconsistent (Cameron & Trivedi, 1998; Fleming & Sorenson, 2004). Poisson models have typically been the first regression attempt when modeling such data. However, since Hausman,

[12] For detailed explanations of the pros and cons behind CEM versus other matching techniques, see Azoulay, Zivin, & Sampat (2010); Iacus, King, & Porro (2011); King (2010); or Blackwell, Iacus, King, & Porro (2002).

Hall, & Griliches (1984) introduced the negative binomial regression, a model which allows for unobserved heterogeneity or overdispersion (conditional variance of data exceeds conditional mean), it has become a favored model in the analysis of patent or publication citation data (Murray & Stern, 2005; Furman & Stern, 2011; Murray, Aghion, Dewatripont, Kolev, & Stern, 2009). Using a dataset composed of citations to articles authored by Chinese (treatment) and US (control) scientists, consider the following conditional fixed effects, negative binomial estimator:

$$Annual\ Citations_{jt} = f(\varepsilon_{jt};\ \gamma_j + \beta_t + \delta_{t\text{-}PubYear} + \Psi{*}China)$$

where γ_j represents article fixed effect and controls for article heterogeneity; β_t represents citation-year fixed effects and controls for changes in citation practices at the time of citation; and $\delta_{t\text{-}PubYear}$ represents article age fixed effects and controls for the nonlinear lifecycle of citations (i.e. the accumulation of citations tends to follow a distinct, non-linear curve (Hall, Jaffe, & Trajtenberg, 2001)). We also run variations of this basic model to control for the number of authors involved and the number of countries represented. The choice of controls for number of involved authors and countries stems from the finding that an increase in team members and diversity tends to increase the number of citations received (Wutchy, Jones, & Uzzi, 2007). Normally a study on citation data would involve some sort of control for subject of the publication; however, since we have already matched on publication issue, we believe that the nature of the research topic has already been sufficiently controlled for prior to the regression. Regression results are displayed in Table 5.

If we constrain the sample of citing publications to journal (document type) articles (publication type), which can be more indicative of real impact to frontier scientific knowledge, the sample of citing publications is reduced to 116,874. As in the previous regression, a number of variations were run with this subset of citing publications. Results are displayed in Table 6.

While the citation count to an article is one measure of impact, another useful consideration is the characteristics of the *citing* publication. For instance, one could count the number of citations to citing publications as another layer of impact measurement. However, gathering this data would be too time intensive, and the usefulness-to-effort ratio would be sharply diminished. Another, simpler measure is the journal impact factor (JIF). As mentioned earlier in this paper, a journal's JIF is calculated as a moving two-year average of the citations received by a paper in that journal. A journal (Journal X) with a JIF equal to five would indicate that an article in Journal X should be expected to receive five citations on the average. The next step of the analysis confines the dependent variable (number of citations) to a pool of high JIF journals[13]. In other words, the relationship of interest is whether or not Chinese-authorship additionally biases articles from receiving citations in *high impact* journals. To increase the sample size of citing publications, we are using all document and publication types. The equation for this model is similar to the previous, with the exception of the change in the dependent variable (regression results displayed in Table 7):

$$Annual\ Citations,\ High\ JIF_{jt} = f(\varepsilon_{jt};\ \gamma_j + \beta_t + \delta_{t\text{-}PubYear} + \Psi{*}China)$$

In order to identify whether or not Chinese authors tend to cite their own, we isolate citations from Chinese-authored publications and run the same regression with this modified dependent variable. We also isolate Asian-authored citing publications (South Korea, Japan, Singapore, and Taiwan), as well as English and American as dependent variables. Regression results are displayed in Table 8.

$$Annual\ Citations,\ Chinese\text{-}Authored_{jt} = f(\varepsilon_{jt};\ \gamma_j + \beta_t + \delta_{t\text{-}PubYear} + \Psi{*}China)$$

The next section of this paper will provide an overview of the regression findings.

IV. Results & Discussion

The coefficient value of negative binomial model, in our case, would be interpreted as follows: the expected number of citations changes by $e^{\mathbf{coefficient}}$ for a Chinese-authored publication over a US-authored one. When interpreting the results of a negative binomial regression, it is typically easier to look at the incidence rate ratio. If the coefficient of the treatment variable (Chinese-authorship) $\beta_{Chinese}$ is equal to $\log(\mu_{Chinese}) - \log(\mu_{US})$, or $\beta_{Chinese} = \log(\mu_{Chinese}/\mu_{US})$, then the incidence rate ratio is $e(\log(\mu_{Chinese}/\mu_{US}))$, or $(\mu_{Chinese}/\mu_{US})$.

In the first regression, which uses as the dependent variable *all citations* (where "all" refers to all document and publication types), Chinese-authorship has a statistically significant and negative impact on the number of citations received to a Nature published article. This effect applies

[13] Note that JIF data is only available for the year 2001-2010, so the sample of data is reduced from earlier regression samples.

robustly to all models, regardless of the control variable specifications, the effects of which were all statistically insignificant. With the exception of model 1 (which excludes age and citation year fixed effects), the incidence rate ratio (IRR) ranges between 0.48 and 0.56 and is significant at 1%. In other words, Chinese first-authorship of Nature derived articles receives an average of half the citation counts as those of US first-authorship.

The second regression, which uses *journal article citations* as the dependent variable, also shows statistical significance in all models for effect of Chinese-authorship. With the exception of model 1, the IRR falls within the range of 0.15 and 0.21 and is significant at 1%. In addition, *number of countries* was statistically significant at 10% in models 3 and 6, and *number of authors* is significant at 10% in model 6. The third regression isolated citations received from *high JIF journals*, and Chinese-authorship was found to be statistically significant in all models at 1%. With the exception of model 1, the IRR ranges between 0.27 and 0.29. In model 3, *number of authors* is statistically significant at 10%. The fourth regression, which isolated citations from *articles published in high JIF journals*, was also statistically significant in the treatment effect at the 1% level for all models. With the exception of model 1, the IRR for Chinese-authorship is approximately 0.06. Control variables were not significant in this regression.

Table 10 shows a comparison of the above four regression models, specifically model 6 from each regression. The highest IRR occurs in the sample comprised of all citations at 0.56, followed by citations from high JIF sources at 0.27, citations from all journal articles at 0.15, and finally citations from high JIF journal articles at 0.06. The IRR of Chinese-authorship in the latter is roughly nine times smaller than that of *all citations*, suggesting that Chinese-authored articles receive an average of one-ninth the citations from *high JIF journal articles* than from the entire pool of citations.

The results from the regressions using country-level subsamples show significance of Chinese-authorship only in the case of the US for models 2 and 6 (other models were not run since the regression coefficients run in a tight range regardless of model choice). The IRR for citations to Chinese papers is in the range of 0.51-0.52 and is statistically significant at 5%. Nature publications first-authored by Chinese scientists appear to have, on average, approximately half the US-derived citations as those authored by US scientists. This own-country preference does not appear to manifest with Asian-originated (Singapore, Japan, South Korea, Taiwan) or Chinese-originated citations. Control variables *number of countries* and the interaction term *number of countries*number of authors* were significant at the 10% level in the Chinese subsample of model 6.

V. Conclusion

The main purpose of this study was to understand whether or not the global scientific community is "standing on the shoulders" of Chinese research. Prior research has established that Chinese-authored papers are, on average, a lower quality than their Western-authored counterparts. It is also possible that Chinese scientists simply are not publishing in areas where citation measures are indicative of impact. To circumvent this selection issue, we isolated Nature and Nature X publications and measured the impact of Chinese-authorship on citation counts for this high quality, quality-controlled sample. Based on the findings, it does not appear that Chinese science is reaching its full potential in terms of impact. In particular, when the sample of citing publications is culled for only high impact journal articles (which are the gold standard in assessing overall impact), we find that Chinese-authored papers receive almost twenty times fewer citations than a similar US-authored paper.

The underutilization of Chinese papers suggests that measures could be taken to improve the efficiency of knowledge diffusion. However, further research should be conducted to understand the exact root of this bias. For instance, exposure to negative press relating cases of academic fraud and "poorly designed" incentives can cause a bias against Chinese-originated science, despite the fact that it is isolated to less prestigious institutions (which are unlikely to publish in journals such as *Nature* or *Science*). Or, perhaps the geographical distance between China and other scientifically advanced countries plays a large role in the citation practices or patterns. It might be the case that social interactions such as attending conferences or, simply, daily casual interaction will increase the likelihood of one scientist impacting another, and thus increase the likelihood of one citing geographically (or institutionally, or culturally) close peers. Our study suggests the existence of a bias against Chinese-authored papers by US-scientists, at least in the citation patterns. However, it does not explain the reasons for this bias. Additionally, the lack of statistical significance in the Asia and China subsamples prevents us from citing the existence of an across-the board, own-country preference in citation patterns. Future research can thus focus on the reasons for this preference and why it appears to exist only in the US (or, additionally, one can study whether this effect exists in other countries outside of the US and Asia).

Ultimately, the goal of this and similar research is to identify areas of inefficiency in the knowledge diffusion process and to develop methods that could be used to improve this process. Now that we have verified the existence of an inefficiency, the next logical step forward is to find its root cause and, finally, to come up with a plan for change and improvement.

VI. Bibliography

Adams, J., & Griliches, Z. (1996). Measuring Science: An Exploration. *NBER Working Paper Series*, *5478*, 1-34.

Aghion, P., & Howitt, P. (1992). A Model of Growth Through Creative Destruction. *Econometrica*, *60* (2), 323-351.

Audretsch, D., & Feldman, M. (1996). R&D spillovers and the geography of innovation and production. *The American Economic Review*, *86* (3), 630-640.

Azoulay, P., Zivin, J., & Sampat, B. (2010, September 3). The Diffusion of Scientific Knowledge Across Time and Space: Evidence from Professional Transitions for the Superstars of Medicine. *Manuscript Prepared for: The Rate and Direction of Inventive Activity, NBER 50th Anniversary Conference*, 1-52.

Bai, L. (2006). Graduate Unemployment: Dilemmas and Challenges in China's Move to Mass Higher Education. *The China Quarterly*, *185*, 128-144.

Blackwell, M., Iacus, S., King, G., & Porro, G. (2009). cem: Coarsened Exact Matching in Stata. *The Stata Journal*, *9* (4), 524-546.

Bonitz, M. (2002). Ranking of nations and heightened competition in Matthew Core journals. *Library Trends*, *50* (3), 440-462.

Bonitz, M. (2005). Ten years Matthew effect for countries. *Scientometrics*, *64* (3), 375-379.

Bonitz, M., & Scharnhorst, A. (1999). Nationa Science Systems and the Matthew Effect for Countries. *Seventh International Conference on Scientometrics and Informetrics.* Colima, Mexico.

Cameron, A., & Trivedi, P. (1998). *Regression Analysis of Count Data.* Cambridge, UK: Cambridge University Press.

Cao, C. (2010, 1 18). *Can Quantity Lead to Quality?* . Retrieved 12 25, 2010, from The New York Times: http://roomfordebate.blogs.nytimes.com/2010/01/18/will-china-achieve-science-supremacy/

Cao, C. (2010, January 26). *Climate for scientific misconduct in China.* Retrieved February 18, 2011, from UPI Asia: http://www.upiasia.com/Society_Culture/2010/01/26/climate_for_scientific_misconduct_in_china/6951/

Chinese Academy of Sciences. (n.d.). Retrieved 12 26, 2010, from http://english.cas.cn/

Connor, S. (2011, February 20). *US science chief warns: 'China will eat our lunch'.* Retrieved August 15, 2011, from The Independent: http://www.independent.co.uk/news/science/us-science-chief-warns-china-will-eat-our-lunch-2219974.html

Dasgupta, P., & David, P. (1994). Toward a new economics of science. *Research Policy , 23*, 487-521.

Dickson, D., & Hepeng, J. (2006, May 31). *China must address the roots of scientific fraud.* Retrieved 2 17, 2011, from Science and Development Network: http://www.scidev.net/en/editorials/china-must-address-the-roots-of-scientific-fraud.html

Elkana, Y., Merton, R. T., & Zuckerman, H. (1978). *Toward a Metric of Science: The Advent of Science Indicators.* New York: Wiley.

Fensterheim, D. (2009). *Knowledge Production at the Global Frontier: The Case of China.* Master's Thesis, MIT, Sloan School of Management, Cambridge, MA.

Fleming, L., & Sorenson, O. (2004). Science as a Map in Technological Search. *Strategic Management Journal , 25* (8/9), 909-928.

Furman, J., & Stern, S. (2011). Climbing atop the Shoulders of Giants: The Impact of Institutions on Cumulative Research. *American Economic Review , 101* (5), 1933-1963.

Garfield, E. (n.d.). *The Thomson Reuters Impact Factor.* Retrieved August 15, 2011, from Thomson Reuters: http://thomsonreuters.com/products_services/science/free/essays/impact_factor/

Hall, B., Jaffe, A., & Trajtenberg, M. (2001, October). The NBER Patent Citations Data File: Lessons, Insights and Methodological Tools (Working Paper No. 8498). *NBER Working Paper* , 1-77.

Harrison, W., Simpson, J., & Weil, M. (2010, January). Editorial. *Acta Crystallographica Section E* , e1-e2.

Hausman, J., Hall, B., & Griliches, Z. (1984). Econometric Models for Count Data with an Application to the Patents-R&D Relationship. *Econometrica , 52* (4), 909-938.

Iacus, S., King, G., & Porro, G. (2011). Causal Inference Without Balance Checking: Coarsened Exact Matching. *Political Analysis* , 1-24.

Jacobs, A. (2010, 10 7). *Rampant Fraud Threat to China's Brisk Ascent.* Retrieved 12 25, 2010, from The New York times: http://www.nytimes.com/2010/10/07/world/asia/07fraud.html?_r=2

Jaffe, A., Trajtenberg, M., & Henderson, R. (1993). Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations. *The Quarterly Journal of Economics , 108* (3), 577-598.

Jha, A. (2011, March 28). *China poised to overhaul US as biggest publisher of scientific papers.* Retrieved August 15, 2011, from The Guardian: http://www.guardian.co.uk/science/2011/mar/28/china-us-publisher-scientific-papers

Kao, J. (2010, January 18). *Will China Achieve Science Supremacy?* Retrieved August 15, 2011, from The New York Times: http://roomfordebate.blogs.nytimes.com/2010/01/18/will-china-achieve-science-supremacy/#cong

King, D. (2004). The scientific impact of nations. *Nature , 430*, 311-316.

King, G. (2010, January 22). Coarsened Exact Matching. *Talk at Washington University, St. Louis* .

Krugman, P. (1991). *Geography and Trade.* Cambridge: MIT Press.

Li, X., & Xiong, L. (1996). Chinese researchers debate rash of plagiarism cases. *Science , 274* (5286), 337-338.

Li, Y., Whalley, J., Shunming, Z., & Xiliang, Z. (2008, April 18). *China's higher education transformation and its global implications.* Retrieved 12 14, 2010, from VOX: http://www.voxeu.org/index.php?q=node/1066

Lim, L. (2011, August 3). *Plagiarism Plague Hinders China's Scientific Ambition.* Retrieved August 15, 2011, from NPR: http://roomfordebate.blogs.nytimes.com/2010/01/18/will-china-achieve-science-supremacy/#cong

Liu, X., & Zhi, T. (2010). China is catching up in science and innovation: the experience of the Chinese Academy of Sciences. *Science and Public Policy , 37* (5), 331-342.

Merton, R. (1957). Priorities in Scientific Discovery: A Chapter in the Sociology of Science. *American Sociological Review , 22* (6), 635-659.

Merton, R. (1942). Science and technology in a democratic order. *Journal of Legal and Political Sociology , 1*, 115-126.

Merton, R. (1968). The Matthew Effect in Science. *Science , 159* (3810), 56-63.

Moed, H., Burger, W., Frankfort, J., & Van Raan, A. (1985). The use of biliometric data for the measurement of university research performance. *Research Policy , 14*, 131-149.

Moed, H., De Bruin, R., & Van Leeuwen, T. N. (1995). New bibliometric tools for the assessment of national research performance: Database description, overview of indicators and first applications. *Scientometrics , 33* (3), 381-422.

Murray, F., & Stern, S. (2005). Do Formal Intellectual Property Rights Hinder the Free Flow of Scientific Knowledge? An Empirical Test of the Anti-Commons Hypothesis. *NBER Academic Science and Entrepreneurship Conference,* (pp. 1-51). Santa Fe, NM.

Murray, F., Aghion, P., Dewatripont, M., Kolev, J., & Stern, S. (2009, May 1). Of Mice and Academics: Examining the Effect of Openness on Innovation. *Working Paper* , 1-37.

Over 10 billion yuan to be invested in "211 Project". (2008, March 26). Retrieved December 25, 2010, from People's Daily Online: http://english.people.com.cn/90001/6381319.html

Romer, P. (1990). Endogenous Technological Change. *The Journal of Political Economy , 98* (5, Part 2: The Problem of Development: A Conference of the Institute for the Study of Free Enterprise Systems), S71-S102.

Royal Society. (2011). *Knowledge, Networks, and Nations: Global scientific collaboration in the 21st century.* London: The Royal Society.

Shukman, D. (2011, March 28). *China to 'overtake US on science' in two years.* Retrieved August 15, 2011, from BBC UK: http://www.bbc.co.uk/news/science-environment-12885271

Sorenson, O., & Fleming, L. (2004). Science and the diffusion of knowledge. *Research Policy , 33* (10), 1615-1634.

Stigler, G. (1979). The Pattern of Citation Practices in Economics. *History of Political Economy , II*, 1-20.

Suttmeier, R. P., Cao, C., & Simon, D. F. (2006, Summer). China's Innovation Challenge and the Remaking of the Chinese Academy of Sciences. *innovations* , 78-97.

The Thomson Reuters Corporation. (2009). *Web of Science 8.0.* Retrieved August 20, 2011, from Thomson Reuters: http://thomsonreuters.com/content/science/pdf/ssr/training/wos_workbook_en.pdf

Thomson Reuters. (2008, July). *Using Bibliometrics: A Guied to Evaluating Research Performance with Citation Data.* Retrieved February 20, 2011, from Thomson Reuters: http://science.thomsonreuters.com/news/2008-07/8465001/

Van Raan, A. (1988). *Handbook of Quantitative Studies of Science and Technology.* Amsterdam: North-Holland Press.

Vinkler, P. (1988). An attempt of surveying and classifying bibliometric indicators for scientometric purposes. *Scientometrics , 13* (5-6), 239-259.

Wang, J. (1999). Chinese Journals Pledge Crackdown. *Science , 283* (5407), 1427.

Wutchy, S., Jones, B., & Uzzi, B. (2007). The Increasing Dominance of Teams in Production of Knowledge. *Science , 316*, 1036-1039.

Xin, H. (2006). Scandals Shake Chinese Science. *Science , 312*, 1464-1466.

Yimin, D. (2001). In China, Publish or Perish is Becoming the New Reality. *Science , 291* (5508), 1477-1479.

VII. Appendix

Nature Auxiliary Journals

Nature & Resources
Nature Genetics
Nature Structural & Molecular Biology
Nature Medicine
Nature and Origin of Amyloid Fibrils
Nature Biotechnology
Nature Neuroscience
Nature Cell Biology
Nature Immunology
Nature Reviews Genetics
Nature Reviews Molecular Cell Biology
Nature Reviews Neuroscience
Nature Reviews Cancer
Nature Reviews Immunology
Nature Materials
Nature Reviews Drug Discovery
Nature Reviews Microbiology
Nature Clinical Practice Cardiovascular Medicine
Nature Clinical Practice Gastroenterology & Hepatology
Nature Clinical Practice Oncology
Nature Clinical Practice Urology
Nature Methods
Nature Chemical Biology
Nature Clinical Practice Endocrinology & Metabolism
Nature Clinical Practice Nephrology
Nature Clinical Practice Neurology
Nature Clinical Practice Rheumatology
Nature Physics
Nature Nanotechnology
Nature Protocols
Nature Photonics

Thomson Reuters Web of Science: List of Variables for Data

File name
Version number
Publication type (e.g. book, journal, book in series)
Authors
Author full names
Group authors
Document title
Full source title
Language
Document type
Author keywords
Keywords Plus ®
Abstract
Author address
Reprint address
E-mail address
Cited references
Cited references count
Times cited
Publisher
Publisher city
Publisher address
ISSN
29-character source abbreviation
ISO source abbreviation
Book series title
Publication year
Publication date
Volume
Issue
Part number
Supplement
Special issue
Beginning page
Ending page
Page count
Article number
Subject category
ISI document delivery number
ISI unique article identifier
End of record
End of file

VIII. Graphs, Tables & Figures

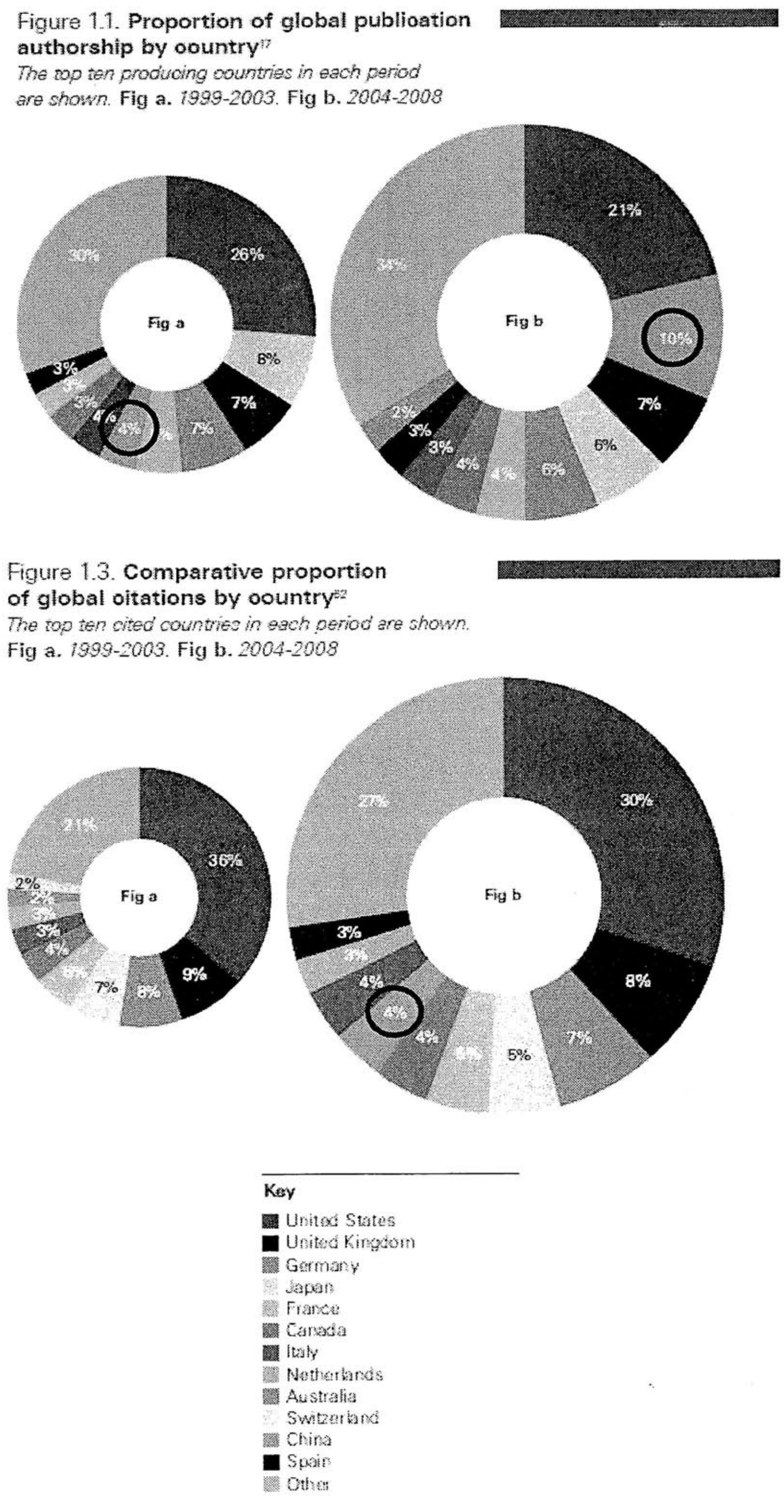

Figure 1: International comparison of growth in publication and citation share
Source: Royal Society

Table 2 **Citation rate per paper normalized to show re-based impact (RBI)**

Country	*2002*	*1993–2002*
Switzerland	1.70	1.59
United States	1.48	1.41
Denmark	1.48	1.33
United Kingdom	1.39	1.21
Netherlands	1.39	1.33
Germany	1.33	1.15
Austria	1.24	1.09
Belgium	1.21	1.17
Sweden	1.21	1.25
Canada	1.18	1.13
Finland	1.18	1.20
France	1.12	1.07
Italy	1.12	1.07
Australia	1.09	1.01
Israel	1.09	1.05
Rep. Ireland	1.03	0.93
Spain	0.97	0.89
Luxembourg	0.94	0.82
Japan	0.91	0.90
Portugal	0.82	0.80
Poland	0.76	0.61
South Africa	0.76	0.61
Greece	0.70	0.67
South Korea	0.64	0.61
Singapore	0.61	0.61
Brazil	0.58	0.62
Russia	0.55	0.40
China	0.55	0.51
Taiwan	0.55	0.56
India	0.48	0.40
Iran	0.42	0.44

1993–2002: average RBI. World RBI = 1
The number of citations, c, to papers in a given discipline for a given year increases with time elapsed since publication, t, approaching an asymptote $c\bullet$ and roughly following an exponential function, $(1 - c/c\bullet) = \exp(-t/t_{1/2})$, where $t_{1/2}$ is a constant, the half-life of this function, which is reached when the citations have reached $c\bullet(1 - 1/e)$. The half-life is shorter in the biological sciences than in the physical sciences. An analysis[2] of papers published in 30 journals in the geological sciences between 1981 and 2002 yields a citation half-life of five years.

Figure 2: Average citations per paper, normalized across subjects
Source: King (2004)

Table 1 Rank order of nations based on share of top 1% of highly cited publications, 1997–2001

	Publications				*Citations*				*Top 1% highly cited publications*			
	1993–97		*1997–2001*		*1993–97*		*1997–2001*		*1993–1997*		*1997–2001*	
Country	*Total*	*Per cent world*	*Total*	*Per cent world*	*Total*	*Per cent world*	*Total*	*Per cent world*	*Total*	*Per cent comparator group*	*Total*	*Per cent comparator group*
United States	1,248,733	37.46	1,265,808	34.86	21,664,121	52.3	10,850,549	49.43	22,710	65.6	23,723	62.76
EU15 (net total)	1,180,730	35.42	1,347,985	37.12	15,147,205	36.57	8,628,152	39.3	11,372	32.85	14,099	37.3
United Kingdom	309,683	9.29	342,535	9.43	4,502,052	10.87	2,500,035	11.39	3,853	11.13	4,831	12.78
Germany	268,393	8.05	318,286	8.76	3,575,143	8.63	2,199,617	10.02	2,974	8.59	3,932	10.4
Japan	289,751	8.69	336,858	9.28	3,123,966	7.54	1,852,271	8.44	2,086	6.03	2,609	6.9
France	203,814	6.11	232,058	6.39	2,638,563	6.37	1,513,090	6.89	2,096	6.05	2,591	6.85
Canada	168,331	5.05	166,216	4.58	2,315,140	5.59	1,164,450	5.3	2,002	5.78	2,195	5.81
Italy	122,398	3.67	147,023	4.05	1,535,208	3.71	964,164	4.39	1,151	3.32	1,630	4.31
Switzerland	57,664	1.73	66,761	1.84	1,113,886	2.69	647,013	2.95	1,196	3.45	1,557	4.12
Netherlands	83,600	2.51	92,526	2.55	1,335,748	3.22	759,027	3.46	1,111	3.21	1,435	3.8
Australia	89,557	2.69	103,300	2.84	1,078,746	2.6	623,636	2.84	852	2.46	1,049	2.78
Sweden	63,757	1.91	72,927	2.01	1,007,418	2.43	548,112	2.5	748	2.16	930	2.46
Spain	79,121	2.37	103,454	2.85	813,722	1.96	559,875	2.55	467	1.35	785	2.08
Belgium	40,147	1.2	48,010	1.32	574,095	1.39	339,895	1.55	482	1.39	639	1.69
Denmark	31,808	0.95	37,198	1.02	508,183	1.23	295,004	1.34	445	1.29	570	1.51
Israel	41,804	1.25	45,944	1.27	517,027	1.25	293,039	1.33	449	1.3	568	1.5
Russia	121,505	3.65	123,629	3.4	509,105	1.23	315,016	1.43	366	1.06	501	1.33
Finland	28,727	0.86	34,690	0.96	427,873	1.03	250,456	1.14	308	0.89	416	1.1
Austria	26,100	0.78	33,598	0.93	332,145	0.8	218,493	1	250	0.72	383	1.01
China	68,661	2.06	115,339	3.18	392,055	0.95	341,519	1.56	153	0.44	375	0.99
South Korea	26,838	0.81	55,739	1.53	183,122	0.44	192,346	0.88	97	0.28	294	0.78
Poland	34,680	1.04	42,852	1.18	237,622	0.57	155,310	0.71	170	0.49	231	0.61
India	72,877	2.19	77,201	2.13	316,461	0.76	188,481	0.86	112	0.32	205	0.54
Brazil	27,874	0.84	43,971	1.21	211,460	0.51	155,357	0.71	100	0.29	188	0.5
Taiwan	32,620	0.98	45,325	1.25	216,852	0.52	150,743	0.69	91	0.26	151	0.4
Rep. Ireland	9,880	0.3	12,779	0.35	104,442	0.25	75,893	0.35	86	0.25	196	0.36
Greece	16,463	0.49	22,333	0.62	128,646	0.31	89,822	0.41	76	0.22	113	0.3
Singapore	9,030	0.27	15,306	0.42	63,288	0.15	55,929	0.25	39	0.11	97	0.26
Portugal	8,102	0.24	13,583	0.37	74,196	0.18	62,814	0.29	43	0.12	96	0.25
South Africa	1,7461	0.52	18,123	0.5	121,598	0.29	67,916	0.31	51	0.15	81	0.21
Iran	2,152	0.06	4,813	0.13	10,706	0.03	12,325	0.06	5	0.01	14	0.04
Luxembourg	300	0.01	430	0.01	2,736	0.01	1,979	0.01	2	0.01	2	0.01
World (net total)	3,333,464	106.23	3,631,368	108.94	41,425,399	118.27	21,953,043	122.97	34,982	127.43	38,263	136.5

This part of the analysis uses a five-year publication window for all disciplines. For papers published 1993–97, the total accumulation of citations to the year 2002 is included. For papers published 1997–2001, the total number of citations to the year 2002 is also included but, given the shorter time period, fewer citations will have accumulated.
The main source of internationally comparable data on research funding, staff and training is the OECD (see 'statistics' at http://www.sourceoecd.org/content/html/index.htm). Data also come from the 2002 editions of the *Main Science and Technology Indicators* and *Basic Science and Technology Statistics*. Accuracy and reliability are discussed in ref. 2. The *Frascati Manual* data definitions and their interpretations of OECD data have been adhered to wherever feasible.

Figure 3: Share of top 1% of highly cited publications, by nation
Source: King (2004)

VARIABLE	DEFINITION
Citation-Year Characteristics	
Forward Citations $_{jt}$	Number of forward citations to Article $_j$ at Year $_t$
Forward Citation, High JIF $_{jt}$	Number of forward citations from high JIF journals to Article j at Year $_t$
Citation Year $_t$	Year in which Forward Citations are received
Age $_{jt}$	Age of article at time of citation (Citation Year $_t$ - Publication Year $_j$)
Root Publication Characteristics	
Publication Year $_j$	Year in which Article $_j$ is published
No. Authors $_j$	Count of the number of authors of Article $_j$
No. Countries $_j$	Count of the number of countries in Article $_j$
Intl. Collaboration $_j$	Dummy variable equal to 1 if more than one country is listed in the address
Chinese Author $_j$	Dummy variable equal to 1 if the first author of Article $_j$ lists a Chinese address
Total Citations $_j$	Number of Forward Citations from Publication Year of Article $_j$ to 2011
Citing Publication Characteristics	
Journal	Dummy variable equal to 1 if the citing journal is from a journal
Article	Dummy variable equal to 1 if the citing journal is an article (vs. Book, Review, Letter, Abstract, etc...)
High JIF	Dummy variable equal to 1 if the citing journal is one of the top 500 most cited at the year of publication
XXX-Authored	Dummy variable equal to 1 if the citing first author lists an address from country XXX
Total Forward Citations	Number of citations received since publication (until 2011) of citing publication

Table 1: Definition of Variables

SAMPLE	VARIABLE	N*	MEAN	ST DEV	MIN	MAX
All	**Citation-Year Characteristics**					
	Forward Citations	9359	18.13	24.98	0	324
	Total Citations	1185	74.58	121.18	1	1445
	Citation Year	9359	2007.23	2.77	2000	2011
	Age	9359	3.77	2.77	0	11
	Article Characteristics					
	Publication Year	1185	2004.10	2.26	2000	2007
	No. Authors	1185	6.88	8.43	1	114
	No. Countries	1185	1.45	0.89	1	8
	Chinese Author	1185	0.11	0.32	0	1
Chinese-Authored (All)	**Citation-Year Characteristics**					
	Forward Citations	1020	13.41	16.28	0	165
	Total Citations	135	45.58	55.44	0	290
	Citation Year	1020	2007.40	2.69	2000	2011
	Age	1020	3.60	3.69	0	11
	Article Characteristics					
	Publication Year	135	2004.44	2.23	2000	2007
	No. Authors	135	9.80	12.12	2	114
	No. Countries	135	1.85	1.05	1	7
US-Authored (All)	**Citation-Year Characteristics**					
	Forward Citations	8339	18.71	25.79	0	324
	Total Citations	1050	78.31	126.73	0	1445
	Citation Year	8339	2007.21	2.78	2000	2011
	Age	8339	3.79	2.78	0	11
	Article Characteristics					
	Publication Year	1050	2004.06	2.26	2000	2007
	No. Authors	1050	6.50	7.75	1	113
	No. Countries	1050	1.40	0.85	1	8

** N varies within samples depending on whether statistics were generated in long versus wide format*

Table 2: Descriptive Statistics, Pre- Match

VARIABLE	N	MEAN	ST DEV	MIN	MAX
Citing Publications					
Journal	169710	0.96	0.19	0	1
Article	169710	0.69	0.46	0	1
Publication Year	169710	2007.26	2.52	2000	2011
Total Forward Citations	168867	27.27	73.44	0	5667
High JIF	169710	0.45	0.50	0	1
No. Countries	169710	1.32	0.78	0	31
Chinese-Authored	169710	0.08	0.26	0	1
Asia-Authored	169710	0.09	0.29	0	1
USA-Authored	169710	0.43	0.50	0	1
England-Authored	169710	0.05	0.23	0	1

Table 3: Descriptive Statistics, Citing Articles

SAMPLE	VARIABLE	N*	MEAN	ST DEV	MIN	MAX
All (Matched)	**Citation-Year Characteristics**					
	Forward Citations	1635	15.96	22.16	0	260
	Total Citations	215	54.57	101.65	0	1238
	Citation Year	1635	2007.26	2.87	2000	2011
	Age	1635	3.74	2.87	0	11
	Article Characteristics					
	Publication Year	215	2004.40	2.57	2000	2007
	No. Authors	215	7.06	4.20	2	30
	No. Countries	215	1.30	0.48	1	3
	Chinese Author	215	0.33	0.47	0	1
Chinese-Authored (Matched)	**Citation-Year Characteristics**					
	Forward Citations	538	14.00	17.92	0	165
	Total Citations	70	46.23	55.22	0	266
	Citation Year	538	2007.24	2.86	2000	2011
	Age	538	3.76	2.86	0	11
	Article Characteristics					
	Publication Year	70	2004.31	2.56	2000	2007
	No. Authors	70	7.71	5.03	2	30
	No. Countries	70	1.39	0.52	1	3
US-Authored (Matched)	**Citation-Year Characteristics**					
	Forward Citations	1097	16.92	23.91	0	260
	Total Citations	145	58.59	117.66	0	1238
	Citation Year	1097	2007.28	2.87	2000	2011
	Age	1097	3.72	2.87	0	11
	Article Characteristics					
	Publication Year	145	2004.43	2.59	2000	2007
	No. Authors	145	6.74	3.72	2	26
	No. Countries	145	1.26	0.46	1	3

** N varies within samples depending on whether statistics were generated in long versus wide format*

Table 4: Descriptive Statistics, Post-Match

	Conditional Fixed Effects Negative Binomial Model Dependent Variable is Annual Citations (All Document and Publication Types) [Incidence-Rate Ratios] Estimated Coefficients (Standard Errors)						
	Model 1	**Model 2**	**Model 3**	**Model 4**	**Model 5**	**Model 6**	**Model 7**
Publication Characteristics							
Chinese Author	**[0.8274] -0.1894 (0.1077)***	**[0.5602] -0.5794 (0.1556)*****	**[0.5603] -0.5793 (0.1558)*****	**[0.5603] -0.5793 (0.1556)*****	**[0.5596] -0.5806 (0.1557)*****	**[0.5567] -0.5857 (0.1562)*****	**[0.4898] -0.7137 (0.1916)*****
Control Variables							
No. Authors	-	-	-	[1.003] 0.0031 (0.0150)	-	[0.9793] -0.0209 (0.0457)	-
No. Countries	-	-	[1.0447] 0.0437 (0.1523)	-	-	[0.8947] -0.1112 (0.3167)	-
Intl. Collaboration	-	-	-	-	-	-	[0.8066] -0.2149 (0.2030)
No. Countries * No. Authors	-	-	-	-	[1.004] 0.0042 (.00923)	[1.0188] 0.0186 (0.0342)	-
Chinese Author * Intl. Collaboration	-	-	-	-	-	-	[1.4777] 0.3905 (0.3288)
Parametric Restrictions							
Age FE	No	Yes	Yes	Yes	Yes	Yes	Yes
Citation Year FE	No	Yes	Yes	Yes	Yes	Yes	Yes
Regression Statistics							
Log-likelihood	-4565.4139	-3883.0311	-3882.9897	-3883.0094	-3882.9267	-3882.8207	-3882.2325
Prob > Chi^2	0.0786	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
No. of Observations	1635	1635	1635	1635	1635	1635	1635

** Significant at 10%, **Significant at 5%, *** Significant at 1%*

Table 5: Regression Results (All Document & Publication Types)

	Conditional Fixed Effects Negative Binomial Model Dependent Variable is Annual Citations (Journal Articles Only) [Incidence-Rate Ratios] Estimated Coefficients (Standard Errors)						
	Model 1	**Model 2**	**Model 3**	**Model 4**	**Model 5**	**Model 6**	**Model 7**
Publication Characteristics							
Chinese Author	**[0.7170] -0.3327 (0.1896)***	**[0.1607] -1.8279 (0.3936)*****	**[0.1582] -1.8438 (0.3894)*****	**[0.1510] -1.890232 (0.3949)*****	**[0.1519] -1.8844 (0.3929)*****	**[0.1499] -1.8978 (0.3910)*****	**[0.2071] -1.5746 (0.3924)*****
Control Variables							
No. Authors	-	-	-	[1.0381] 0.0374 (0.0244)	-	**[1.1312] 0.1234 (0.0741)***	-
No. Countries	-	-	**[1.5321] 0.4266 (0.2581)***	-	-	**[2.4020] 0.8763 (0.4584)***	-
Intl. Collaboration	-	-	-	-	-	-	[1214171] 14.0010 (520.5754)
No. Countries * No. Authors	-	-	-	-	[1.0229] 0.0226 (0.0153)	[0.9363] -0.0658 (0.0480)	-
Chinese Author * Intl. Collaboration	-	-	-	-	-	-	[9.6e-07] -13.8517 (520.5755)
Parametric Restrictions							
Age FE	No	Yes	Yes	Yes	Yes	Yes	Yes
Citation Year FE	No	Yes	Yes	Yes	Yes	Yes	Yes
Regression Statistics							
Log-likelihood	-1947.3239	-1635.7652	-1634.3392	-1634.448	-1634.4426	-1632.4833	-1632.6068
Prob > Chi^2	0.0794	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
No. of Observations	716	716	716	716	716	716	716

** Significant at 10%, **Significant at 5%, *** Significant at 1%*

Table 6: Regression Results (Journal Articles Only)

	Conditional Fixed Effects Negative Binomial Model Dependent Variable is Annual Citations (High JIF Citations) [Incidence-Rate Ratios] Estimated Coefficients (Standard Errors)						
	Model 1	**Model 2**	**Model 3**	**Model 4**	**Model 5**	**Model 6**	**Model 7**
Publication Characteristics							
Chinese Author	**[0.4587] -0.7793 (0.1251)*****	**[0.2699] -1.3097 (0.1780)*****	**[0.2705] -1.3073 (0.1776)*****	**[0.2704] -1.3080 (0.1792)*****	**[0.2689] -1.3134 (0.1788)*****	**[0.2709] -1.3060 (0.1787)*****	**[0.2911] -1.2342 (0.2238)*****
Control Variables							
No. Authors	-	-	-	**[1.0460] 0.0450 (0.0244)***	-	[1.0765] 0.0737 (0.0726)	-
No. Countries	-	-	[0.8770] -0.1312 (0.1571)	-	-	[0.9871] -0.0130 (0.3872)	-
Intl. Collaboration	-	-	-	-	-	-	[0.9559] -0.0451 (0.2354)
No. Countries * No. Authors	-	-	-	-	[1.0115] 0.0115 (0.0123)	[0.9805] -0.0197 (0.0488)	-
Chinese Author * Intl. Collaboration	-	-	-	-	-	-	[0.8233] -0.1944 (0.3579)
Parametric Restrictions							
Age FE	No	Yes	Yes	Yes	Yes	Yes	Yes
Citation Year FE	No	Yes	Yes	Yes	Yes	Yes	Yes
Regression Statistics							
Log-likelihood	-3407.6162	-3011.3859	-3011.0425	-3009.3107	-3010.9135	-3008.7596	-3008.7596
Prob > Chi^2	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
No. of Observations	1629	1629	1629	1629	1629	1629	1629

** Significant at 10%, **Significant at 5%, *** Significant at 1%*

Table 7: Regression Results (All Citations, High JIF)

	Conditional Fixed Effects Negative Binomial Model Dependent Variable is Annual Citations (High JIF Citations from Journal Articles) [Incidence-Rate Ratios] Estimated Coefficients (Standard Errors)						
	Model 1	**Model 2**	**Model 3**	**Model 4**	**Model 5**	**Model 6**	**Model 7**
Publication Characteristics							
Chinese Author	**[0.1433] -1.9426 (0.1386)*****	**[0.0621] -2.7791 (.01937)*****	**[0.0625] -2.7718 (0.1943)*****	**[0.0621] -2.7788 (0.1936)*****	**[0.0625] -2.7729 (0.1939)*****	**[0.0622] -2.7774 (0.1946)*****	**[0.0643] -2.7443 (0.2520)*****
Control Variables							
No. Authors	-	-	-	[0.9877] -0.0124 (0.0197)	-	[0.9592] -0.0416 (0.0587)	-
No. Countries	-	-	[0.9302] -0.0723 (0.1684]	-	-	[0.7960] -0.2282 (0.3529)	-
Intl. Collaboration	-	-	-	-	-	-	[0.9821] -0.0181 (0.2666_
No. Countries * No. Authors	-	-	-	-	[0.9941] -0.0059 (0.0108)	[1.0215] 0.0213 (0.0397)	-
Chinese Author * Intl. Collaboration	-	-	-	-	-	-	[0.9288] -0.0738 (0.3932)
Parametric Restrictions							
Age FE	No	Yes	Yes	Yes	Yes	Yes	Yes
Citation Year FE	No	Yes	Yes	Yes	Yes	Yes	Yes
Regression Statistics							
Log-likelihood	-2558.9666	-2160.4467	-2160.3551	-2160.2528	-2160.2972	-2160.0414	-2160.3944
Prob > Chi^2	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
No. of Observations	1589	1589	1589	1589	1589	1589	1589

** Significant at 10%, **Significant at 5%, *** Significant at 1%*

Table 8: Regression Results (Journal Article Citations, High JIF)

	Model 2			Model 6		
	Asia	USA	China	Asia	USA	China
Publication Characteristics						
Chinese Author	[0.5161] -0.6615 (0.6676)	**[0.5224]** **-0.6493** **(0.3032)****	[0.4842] -0.7252 (0.9251)	[0.5239] -0.6464 (0.6641)	**[0.5109]** **-0.6716** **(0.2993)****	[0.6058] -0.5012 (0.7768)
Control Variables						
No. Authors	-	-	-	[0.8891] -0.1176 (0.4047)	[1.1677] 0.1550 (0.0991)	[1.5330] 0.4272 (0.2664)
No. Countries	-	-	-	[0.6045] -0.5034 (2.4743)	[1.9421] 0.6638 (0.6718)	**[4.1371]** **1.4200** **(0.8372)***
Intl. Collaboration	-	-	-	-	-	-
No. Countries * No. Authors	-	-	-	[1.1827] 0.1678 (0.3715)	[0.9047] -0.1002 (0.0744)	**[0.7913]** **-0.2340** **(0.1389)***
Chinese Author * Intl. Collaboration	-	-	-	-	-	-
Parametric Restrictions						
Age FE	Yes	Yes	Yes	Yes	Yes	Yes
Citation Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Regression Statistics						
Log-likelihood	-1271.8828	-2859.3266	-1440.4309	-1271.0224	-2857.6895	-1436.6212
Prob > Chi^2	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
No. of Observations	1417	1623	1355	1417	1623	1355

Conditional Fixed Effects Negative Binomial Model
Dependent Variable is Annual Citations from Asia, USA, or China
[Incidence-Rate Ratios]
Estimated Coefficients
(Standard Errors)

** Significant at 10%, **Significant at 5%, *** Significant at 1%*

Table 9: Regression Results (Citations by Country/Region)

	Conditional Fixed Effects Negative Binomial Model Regression Model 6: Comparison [Incidence-Rate Ratios] Estimated Coefficients (Standard Errors)			
	All	**Journal Article**	**All High JIF**	**High JIF Journal Articles**
Publication Characteristics				
Chinese Author	**[0.5567] -0.5857 (0.1562)*** **	**[0.1499] -1.8978 (0.3910)*** **	**[0.2709] -1.3060 (0.1787)*** **	**[0.0622] -2.7774 (0.1946)*** **
Control Variables				
No. Authors	[0.9793] -0.0209 (0.0457)	[1.1312] 0.1234 (0.0741)*	[1.0765] 0.0737 (0.0726)	[0.9592] -0.0416 (0.0587)
No. Countries	[0.8947] -0.1112 (0.3167)	[2.4020] 0.8763 (0.4584)*	[0.9871] -0.0130 (0.3872)	[0.7960] -0.2282 (0.3529)
Intl. Collaboration	-	-	-	-
No. Countries * No. Authors	[1.0188] 0.0186 (0.0342)	[0.9363] -0.0658 (0.0480)	[0.9805] -0.0197 (0.0488)	[1.0215] 0.0213 (0.0397)
Chinese Author * Intl. Collaboration	-	-	-	-
Parametric Restrictions				
Age FE	Yes	Yes	Yes	Yes
Citation Year FE	Yes	Yes	Yes	Yes
Regression Statistics				
Log-likelihood	-3882.8207	-1632.4833	-3008.7596	-2160.0414
Prob > Chi^2	0.0000	0.0000	0.0000	0.0000
No. of Observations	1635	716	1629	1589

** Significant at 10%, **Significant at 5%, *** Significant at 1%*

Table 10: Comparison of Model 6