Who are the Best and Brightest Researchers in the U.S.?

Wei Huang*

Abstract

Using a unique dataset of scientific journal articles with US-based authors, I compare scientific productivity of non-Anglo-Saxon researchers with that of Anglo-Saxon ones. Estimates show that papers with first author having a non-Anglo-Saxon name have higher impact factor (IF) and more citations, which can be largely explained by organization and last author fixed effects. However, a contrary pattern exists for the last author. To explain the phenomenon, I examine and provide evidence for that relatively more talented Anglo-Saxon first authors will become last authors in US (selection effect) and Anglo-Saxon last authors have more experience in academia (cohort effect). In addition, I also find that the quality of papers written by non-Anglo-Saxon researchers has been consistently increasing since late 1980s relative to those by Anglo-Saxon ones, and this is mainly because US is attracting researchers with relatively higher ability. Finally, I find that non-Anglo-Saxon authors write more papers than Anglo-Saxon ones do in total.

JEL Codes: F22, I23, J61, O15, O33

Keywords: Non-Anglo-Saxon, Impact factor, Citations

(First draft for comments only)

* Department of Economics, Harvard University. Email: weihuang@fas.harvard.edu. I thank my advisors Richard Freeman and Laurence Katz for their helpful suggestions and persistent guidance. I also thank the comments from Professor Amanda Pallais. All errors are mine.

Introduction

The globalization of science has changed the ethnic and national origin of scientists and engineers in the US. Over half of post-doctorate workers in science labs come from overseas. Expansion of higher education worldwide has increased the supply of non-US educated scientific researchers, which contributes to the flow of immigrant scientists and engineers to the US. Political shocks such as the collapse of the Soviet Union have led to sudden influxes of scientists and engineers to the US job market.[1] The share of the foreign-born among science and engineering researchers has been raised by not only the lasting influx of international students and post-doctorate workers (Freeman, 2009; Freeman and Huang, 2012). It is natural to ask how these immigrant researchers perform in US relative to native ones.

Answering the above question is important, both for academic research and policy implication as well. Earlier research on immigration and scientific productivity has focused on the role of foreign-born star scientists in US Science and on the propensity of natives and migrants to publish during their career. Levin and Stephan (1999) and Hunt (2009) show that foreign-born US scientists tend to outperform natives on a number of measures of scientific productivity. Gaule and Piacentini (2012) use data from Proqueset Dissertation and Abstracts to identify the positive effect of Chinese Chemistry students on scientific productivity and find strong evidence for that.

This study tries to shed some light on the question above by analyzing the scientific outcomes of researchers from different ethnic background. I mainly focus on the comparison between Anglo-Saxon authors and non-Anglo-Saxon ones because the latter is mainly composed of immigrants. Using a unique dataset about the published scientific papers from Web of Science and matching the likely ethnicities by authors' last names, I estimate the outcomes of the non-Anglo-Saxon authors relative to those of Anglo-Saxon ones, which is measured by the impact factor of the journal where the article is published, citation rank constructed from the number of being cited till 2009 and number of papers written by the author, respectively.

[1] John Bound, Sarah Turner, Patrick Walsh, "Internationalization of U.S. Doctorate Education" in Science and Engineering Careers in the United States: An Analysis of Markets and Employment (2009), Richard B. Freeman and Daniel L. Goroff, editors (p. 59 – 97). George J. Borjas, Kirk B. Doran "The Collapse of the Soviet Union and the Productivity of American Mathematicians" The Quarterly Journal of Economics, 2012.

Our analysis shows that papers with first authors being non-Anglo-Saxon are associated with remarkably higher quality, which can be largely explained by the organization where the authors do the research, and the coauthors with whom the researchers work. In stark contrast to this, however, last authors being non-Anglo-Saxon is negatively associated with the paper quality, and neither of two factors could explain it at all.

Different positions in scientific papers reflect the authors' identity. In most scientific fields, the first-author is the junior researcher who did the most work on the paper while the last author is the senior person in whose laboratory the work was conducted.[2] Data from Web of science support this claim because I find that, relative to first authors, the last authors are more likely to have written at least one paper and average impact factor of pervious papers is also higher. Based on this fact, I test the hypotheses that relatively more talented Anglo-Saxon first authors will become last authors in US (selection effect), and that Anglo-Saxon last authors have more experience in academia (cohort effect). Estimates in this study support both of these hypotheses but cannot quantify how much they are.

Furthermore, results also indicate that the papers written by non-Anglo-Saxon authors have a constant increase in quality since late 1980s, and further analysis reveals the main driving force is the entry of newcomers with higher ability or talent, indicating that US has been actually attracting high-quality researchers since late 1980s.

Because it is possible that researchers only write higher quality papers at the cost of quantity, do non-Anglo-Saxon first authors write fewer papers during this time? I find that the non-Anglo-Saxon authors, no matter in the first or the last position, do write more papers not only in total during this period but also in a single active year (when the authors write at least one paper).

There is still one concern because non-Anglo-Saxon authors are not equivalent to immigrants. I then differentiate the Chinese into the American-born and China-born by exploiting their first name initials, because China-born Chinese are mostly immigrants but the American-born Chinese are usually native speakers. The analysis shows that the first author premium is mainly contributed by the China born Chinese and that Chine born Chinese outperform the American born, measured by paper quality. These results are consistent with positive selection theory and

[2] For example, if the first author is a Ph.D student, then the last author is likely to be the academic advisor. While intermediate positions reflect contributions of researchers who contributed in other ways.

also emphasis the role of immigrants in science research. (Borjas, 1987; Jasso and Rosenzweig, 1990; Gaule and Piacentini, 2012)

The paper is organized as follows. Section 2 introduces data I used in this study and provides summary statistics. Section 3 shows empirical results and Section 4 concludes.

I. Data

2.1 Ethnic Composition of US-based Researchers

Using data from Thomson Reuters Web of Science, I created a database of US-based academic two-, three- and four- authored articles published between 1985 and 2007. We limited our sample to US-based authors because US is the hosting country in this analysis, where produce over 30 percent English articles all over the world. I excluded social science papers (0.3% in total) and limited the sample to authors of 2-4 natural scientific academic articles so that each paper have different first author and last author, and these papers captures about half coauthored papers.[3] The data set provides authors' complete surnames, initials of first names, addresses, citation and impact factor.[4]

I then applied a program developed by Bill Kerr that matches the names of persons to their likely ethnicity. This program uses names and MSAs to determine the likely ethnicity of authors. Names such as Kim are far more likely to represent Korean people than any other, while names like Zhang are likely to be Chinese. Because persons of a particular ethnicity are more likely to live in some MSAs than others, MSA information helps distinguish ethnicity among people as well.[5] [6] I matched surnames to ethnicity at a rate of 80%. This is lower than that in usual matching with both given names and surnames (about 95%). I restrict the sample used to the papers with first and last authors identified to belong to one or some ethnicities, which contains

[3] Richard Freeman, Ina Ganguli, and Raviv Goroff-Muriciano "International Collaboration in Research" (The Changing Frontier: Rethinking Science and Innovation Policy, Pre-Conference, NBER, Oct 26, 2012) find that 55% of coauthored papers in nano-technology, 58% of coauthored papers in particle physics and 52% of coauthored papers in biotechnology and applied microbiology have 2, 3, or 4 authors.

[4] Since some authors may have multiple addresses and different authors may have different ones, we are not able to identify the address for a specific author in a paper because the addresses of the authors are pooled together.

[5] Ethnicity is divided into nine categories: Chinese (CHN), Anglo-Saxon/English (ENG), European (EUR), Indian/Hindi/South Asian (HIN), Hispanic/Filipino (HIS), Japanese (JAP), Korean (KOR), Russian (RUS) and Vietnamese (VNM).

[6] More details can be found in William R. Kerr and William F. Lincoln, "The Supply Side of Innovation: H-1B Visa Reforms and US Ethnic Invention," The Journal of Labor Economics 28:3 (July 2010), 473-508. and William R. Kerr, "Ethnic Scientific Communities and International Technology Diffusion," The Review of Economics and Statistics, 90:3 (August 2008), 518-537.

1,164,919 papers in 11 fields.[7] The number of papers increases by 112 percent from 33,789 in 1985 to 71,531 in 2008. Panel A in Table A1 reports the proportions of non-Anglo-Saxon names in first authors and last authors of these papers, which are 0.50 and 0.38, respectively. The proportions are very consistent in two-, three- and four-authored papers.

[Table A1 about here]

Figure 1 plots the proportion of Non-Anglo-Saxon names in the sample against publication year, by author position. From 1985 to 2007, the proportion of Non-Anglo-Saxon first authors increases by 42 percent from 39 percent to 56 percent. The pattern in last author is similar but the speed of increasing is slower: the proportion of Non-Anglo-Saxon names increases from 32 to 44 percent. Statistics for first and last authors by Non-Anglo-Saxon ethnicities are plotted in Figure A1a and A1b, respectively. What is featured in these figures is a dramatic increase of proportion of Chinese authors: from 1985 to 2007, the proportion of Chinese first authors triples from 6.7 to 20.1 percent, and that of Chinese last authors increases from 4.9 to 11. During the same period, however, the proportion of European names has been decreasing, no matter first author or last author position.

It should be kept in mind that Non-Anglo-Saxon names are not equivalent to immigrants. American Born Chinese, for example, are not immigrants, though they are likely to have a Chinese family name. Following Kerr and Lincoln (2010), I do not differentiate the non-Anglo-Saxon authors and immigrants first because 1) majority of non-Anglo-Saxon authors are immigrants,[8] and 2) the differences between Anglo-Saxon authors and non-Anglo-Saxon ones may be reasonable to be recognized as long-term effects because the later group includes second and further immigrant generations.[9] However, I still distinguish the China Born Chinese (CBC) and American Born Chinese (ABC) by using the different distribution of first name initials under the assumption that ABC are given an American Name while CBC usually have traditional Chinese names. More details can be found in Section 3.

2.2 Article information

[7] These fields are Multidisciplinary (2%), Agriculture (11%), Biology (7%), Biomedicine (26%), Chemistry (11%), Clinical Medicine (31%), Engineering (8%), Geosciences (4%), Information Computer Technology (6%), Material Science (3%), Mathematics (3%) and Physics (10%).

[8] As estimated in Section 3, there are over 70 percent of Chinese authors are native Chinese.

[9] As long as family names keep the same.

In this study, I use two different paper quality measures: impact factor and citation rank. The impact factor of a Web of Science journal in a year is the average number of citations to articles in the journal in the preceding two years.

For each publication year, I define the citation rank of a paper with c times of citations as $\frac{N_{citation<c}+1}{N_{total}}\times 100$, in which $N_{citation<c}$ is the number of papers with citations less than c, and N_{total} is the total number of papers published in the same year. The reason why I do not use number of citations directly is that it follows power law distribution (Redner, 2005; Gupta et al., 2005), which would bring potential econometric problems if used directly in linear regression models. This citation rank ranges from nearly zero to one hundred. For a particular paper, it can be understood as the proportion of articles published in the same year that have fewer citations, in percentage terms.

Panel B in Table A1 reports statistics for the article information, including impact factor and citation rank, as well as number of references and number of addresses. The mean values of impact factor of two-, three- and four-authored papers are 2.31, 2.60 and 2.93, respectively, and those of citation rank are 44.21, 48.21 and 52.20. Similarly, papers with more authors also have more references and addresses.

2.3 Author Identifier and Distinction between First and Last Authors

Web of Science Data provide full last name and first name initials of all the authors. Unfortunately, different authors may share the same names and that will be a problem in identification. Our sample contains over 1.16 million papers and over 3.32 million names. Dividing names by papers, we have an average of 2.86 authors per paper. But many of the names in the data set are the same. The disambiguation problem is to differentiate which of the multiple appearances of the names reflect the same person writing more than one paper and which reflect different persons with the same name writing some of those papers. Following the same methodology in Freeman and Huang (2013), I differentiate the same name into separate people by using the 12 fields under the assumption that people with the same names writing in different fields are in fact different. This yields a total number of identified distinct first authors of 560,771 and 562,560 last authors in the sample. However, it is noteworthy that these authors may not be exactly distinct either, since the possibility also exists that some authors in the same fields have same names.

In most scientific fields, the first-author is the *junior* researcher who did the most work on the paper while the last author is the *senior* person in whose laboratory the work was conducted. Using identifiers of first and last authors defined in last paragraph, I provide evidence for this claim by reporting the average impact factor and citation rank of previous papers written by the same person (identified by the identifiers) in Panel C of Table A1. It is obvious that last authors write not only more papers but also better ones (measured by impact factor). The differences are large and significant.

II. Empirical Results

3.1 Non-Anglo-Saxon Names and Paper Quality

3.1.1 Paper Quality of Non-Anglo-Saxon Names, by Author Position

I assess the paper quality difference between Anglo-Saxon authors and the others by estimating the following linear equation:

$$Y_{i,t} = \beta_0 + \beta_f Eth_f + \beta_l Eth_l + \gamma X_{i,t} + \varepsilon_i \qquad (1)$$

in which the dependent variable $Y_{i,t}$ denotes the paper quality of paper i published in year t (i.e. impact factor or citation rank), and Eth_f and Eth_l are the probability that first or last author has non-Anglo-Saxon name, respectively, which ranges from zero to one. The coefficients, β_f and β_l, are of central interest in this study, because they reflect the quality difference between papers written by Anglo-Saxon authors and those by non-Anglo-Saxon ones. $X_{i,t}$ denotes a set of covariates, including number of addresses, number of references, dummies for number of authors, dummies for publication years, dummies for 50 states and 180 subfields based on the journal of publication as well. All these variables are controlled for because they are potentially correlated with both authors' ethnicity probability and paper quality. For example, non-Anglo-Saxon authors may be concentrated in certain scientific fields[10] and papers in these fields may have relatively higher/lower impact factor or more/less citations.

Estimates for β_f and β_l in Equation (1) and its extensions are presented in Table 1. Dependent variable is impact factor in Panel A and citation rank in Panel B. Estimates in column 1 show that papers whose first author is non-Anglo-Saxon have relatively 0.07 higher impact

[10] The fields include Material Science, Physics, ICT, etc. For example, the proportions of non-Anglo-Saxon first authors in these fields are 0.64, 0.63 and 0.58, respectively. But those in agriculture and geosciences are 0.39 and 0.41.

factor (2.8 percent of the mean) and 0.78 higher citation rank (1.7 percent of the mean).[11] Both of them are significant at 1 percent significance level. However, last author being non-Anglo-Saxon is significantly associated with lower impact factor, and the sign of the corresponding variable is also negative when dependent variable is citation rank, though it is not significant.[12]

[Table 1 about here]

Is it possible that non-Anglo-Saxon first authors are disproportionally concentrated in higher ranked institutions and those non-Anglo-Saxon last authors in relatively lower ones so that non-Anglo-Saxon first authors have quality-premium and non-Anglo-Saxon last authors have quality-discount? Column 2 tests this hypothesis by controlling for the organization fixed effects. Comparing to the coefficients in column 1, I find that the non-Anglo-Saxon first author premium can be explained a lot (37 percent for impact factor and 59 percent for citation rank). However, the non-Anglo-Saxon last author discount can hardly be explained by the organization fixed effect: the magnitudes of the coefficients become larger and more significant. These results possibly reflect that non-Anglo-Saxon first and last authors are more likely to be in better institutions relative to Anglo-Saxon ones.

Coauthorship also plays an important role in paper quality (Freeman and Huang, 2013), and columns 3 throughout 5 examine whether the coauthorship accounts for the first author premium and last author discount. Specifically, if the non-Anglo-Saxon first authors are more likely to coauthor with more talent last authors and it is opposite for non-Anglo-Saxon last authors, then the premium or discount can possibly explained. Results in column 3 where last author identifier fixed effects are controlled for, show that the first author premium can be largely explained (50 percent for impact factor and 97 percent for citation rank). When dummies for organization and last author identifier and their interactions are added, the first author premium can be further explained (64 percent for impact factor and 100 percent for citation rank).[13] I control for the first author identifier fixed effects in column 5, only to find that the last author discount still exists.

Thus, I conclude from Table 1 that 1) first authors being non-Anglo-Saxon is positively associated with higher paper quality, but opposite for last authors, and 2) organization fixed effects and coauthorship account for the first author premium largely but hardly explain the last author discount.

[11] I name it as first author premium hereafter.
[12] I name it as last author discount hereafter.
[13] The coefficient for citation rank is negative and insignificant. I calculate it as zero.

I further divide the non-Anglo-Saxon authors into eight different ethnicity categories as mentioned above, conduct the same regressions, and report the results in Table A2. The estimates show that there exists significant heterogeneity across different ethnicities, but the general pattern is consistent with that in Table 1.

[Table A2 about here]

In addition, I also provide the results for different fields in Table A3a, and show the proportions of non-Anglo-Saxon names in different fields by author position in Table A3b. These results show large heterogeneity across different fields, both in the paper quality and in the distribution of non-Anglo-Saxon names.

[Table A3a and A3b about here]

3.1.2 Selection Effect and Cohort Effect

This section answers the question why non-Anglo-Saxon first authors write higher quality paper but those last authors follow the opposite pattern. Two obvious hypotheses are 1) relatively more talent Anglo-Saxon first authors (junior academics) become last authors (senior academics), which is *selection effect*,[14] and 2) Anglo-Saxon last authors are relatively older than non-Anglo-Saxon ones since the large proportion of the later group are new comers, and the last author discount thus reflects the difference in experience, which is *cohort effect*.[15]

For simplicity, I assume that established seniors are always last authors and juniors are always first authors until they become seniors. Then, I test the two hypotheses in the following procedure. First, match the identifiers of first authors of earlier year papers to those of last authors of later year papers. Second, divide the whole sample by whether the last author is matched or not and whether the first author is matched or not, respectively.[16] Third, estimate Equation (1) in the four different subsamples. On one hand, if the hypothesis of *cohort effect* is true, the coefficient on last author being non-Anglo-Saxon in the sample where last authors are matched, β_l^M, is larger than that in the sample where last authors are not matched, β_l^{NM}, (i.e. $\beta_l^M > \beta_l^{NM}$), because the established seniors are in the "not matched" sample and Anglo-Saxon

[14] Immigrants potentially choose whether to stay in US, and whether to pursue career in academy (Grogger and Hanson, 2013).

[15] Unfortunately, WOS do not provide information on authors' age or experience.

[16] I trim the sample to those papers published between 1988 and 2005 because the last authors of the earliest years (1985-1987) and the first authors of the latest years (2006-2007) can hardly to be matched. But the results are consistent when papers in these years are included.

seniors outperform their non-Anglo-Saxon counterpart. On the other hand, if the hypothesis of *selection effect* holds, the coefficient on first author being non-Anglo-Saxon in the sample where first authors are matched, β_f^M, should be smaller than that in the sample where first authors are not matched, β_f^{NM}, (i.e. $\beta_f^M < \beta_f^{NM}$), because only Anglo-Saxon first authors with relatively higher talent are selected into the "matched" sample.

Table 2 reports the estimation results. First five columns show the results with basic covariates controlled for, including number of addresses, dummies for publish years, number of authors, states and subfields, and the results in rest five columns also include organization fixed effects. The coefficients mentioned in last paragraph are marked under grey shadow. Dependent variable is impact factor in Panel A, and citation rank in Panel B. First of all, I report in column 1 and 6 to show that the first author premium and last author discount also exist in this trimmed sample, which are consistent with the results in first two columns of Table 1. What is more important, the coefficients in grey areas provide evidence for both *cohort effect* and *selection effect* hypotheses. Just take the columns 7 through 10 in Panel B for an example. The β_l^M is estimated to be -0.175 (0.121) and β_l^{NM} is -0.619 (0.114), indicating *cohort effect* exists provide that $\beta_l^M > \beta_l^{NM}$. Similarly, β_f^M is estimated to be -0.073 (0.135), which is smaller than estimated β_f^{NM}, 0.722 (0.100), and this is consistent with prediction of the *selection effect* hypothesis. All the other coefficients in grey areas also provide consistent pattern, though some of the differences are not significant.

[Table 2 about here]

Table A3 also presents the results when dividing the non-Anglo-Saxon group into the eight different ethnicities, and they are generally consistent with those in Table 2 provided that at least four out of eight are consistent with the two hypotheses in each grey area. One concern refers to the name disambiguation problem that the same identifiers may correspond to different individuals. It is possible that those authors who are matched in different years are actually distinct researchers. It potentially matters because the papers of these *wrongly* matched authors will be included in the "matched" sample. However, this may be not a huge problem because the proportion of "popular" names is small: less than 9 percent of the authors' identifiers appear in over five years. In addition, though the extent of name disambiguation varies across different ethnicities, results in Table A4 still show a consistent pattern in general. Particularly, concern may rise provided that more Chinese may have the same last names relative to individuals of other ethnicities, so it is easier for Chinese names to be wrongly matched. It is important given that Chinese compose a large proportion of non-Anglo-Saxon authors. Comparing the coefficients in Table A4 and Table 2, I find that the some evidence still holds for both *selection effect*[17] and *cohort effect*.[18]

[Table A4 about here]

3.1.3 *Non-Anglo-Saxon Names and Paper Quality in different periods*

Above analysis treat the whole sample as static pool, but it is natural to ask how the quality difference between papers by Anglo-Saxon and those by non-Anglo-Saxon changes over time. Since the proportion of papers written by non-Anglo-Saxon increase remarkably in the past decades and the immigrants may contribute most of them, answering the above question will shed some light on issues on current development of science research and immigration policy.

I examine it by dividing the full sample into four distinct subsamples by publication year (1985-1990, 1991-1995, 1996-2000 and 2001-2007) and estimating Equation (1) separately. Table 3 reports the results. The first four columns include basic covariates and the last four add organization fixed effects. I find that the non-Anglo-Saxon authors have been doing better and better over time (relative to Anglo-Saxon authors).[19] Among the papers published in the earliest period (1985-1990), the ones written by non-Anglo-Saxon authors, no matter first or last position, have lower impact factor and fewer citations. To the contrary, the latest ones (published in 2001-2007) with non-Anglo-Saxon first authors now receive higher impact factor and more citations, and the ones with non-Anglo-Saxon last authors also have more citations.

[Table 3 about here]

But why are the non-Anglo-Saxon authors doing better over time? Is it because the ones who came earlier improve more during their stay (Stayer hypothesis) or U.S. attract more talented

[17] Papers written by Chinese first authors have relatively higher impact factor and more citations among non-Anglo-Saxon first authors. If the wrongly matched rate for Chinese is disproportionally high, the estimated coefficients β_f^M would be upward biased and estimated β_f^{NM} would be downward biased. Even though, estimates in Table 2 still consistently show that $\beta_f^M < \beta_f^{NM}$, indicating *selection effect* still exists.

[18] By comparing the results in Table A3 and those in columns 1 and 6 in Table 2, I find that Chinese last authors write papers with lower impact factor but more citations among non-Anglo-Saxon authors. When dependent variable is impact factor, this higher wrongly matching rate in Chinese will bias the β_l^M downward. But Panel A of Table 2 still shows $\beta_l^M > \beta_l^{NM}$.

[19] Except for the non-Anglo-Saxon last authors from 1986-1990 to 1991-1995.

researchers from other countries (Newcomer hypothesis)?[20] To test the latter hypothesis is meaningful because immigration policy makers may be interested whether U.S. current policy does attract the talent ones to come and work in U.S., and my empirical results provide a positive response.

I do the following to test them. In each subsample by publication year period, I further divide the papers by whether the first author identifier appear in first author position in previous papers, and whether the last author identifiers appear in last author position in previous papers, respectively. Then I estimate Equation (1) in each subsample.[21] On one hand, I compare the coefficients on non-Anglo-Saxon authors in the "new name" sample to that in the "older name" sample of the same period. If newcomer hypothesis is true, the former coefficient should be larger. On the other, I also compare the coefficients on non-Anglo-Saxon authors in the "new name" sample to that in the "older name" sample of the next period, to see the track of difference between Anglo-Saxon and non-Anglo-Saxon over time with the assumption that the "new" names authors will become "older" name ones in the next period.

Table 4 reports the results. The first four columns show the results when dependent variable is impact factor and the rest four when that is citation rank.[22] Different panels show the results for papers published in different periods. The coefficients in grey areas are of interest. Consistent with Newcomer hypothesis, the coefficient in "new" name sample is larger than that in "older" name sample of the same period in each grey area. Comparing the coefficient in new name sample to that in older name sample of next period, I do not find significant change for first author and a small drop for last author.[23] Therefore, these results provide evidence for Newcomer hypothesis but reject the Stayer hypothesis. So I conclude here that the relative increasing paper quality of non-Anglo-Saxon authors is because U.S. receives newcomers with higher ability or talent.

[Table 4 about here]

3.3 Non-Anglo-Saxon Names and Paper Quantity

The analysis above mainly focuses on paper quality, and I find papers written by non-Anglo-

[20] Relative to Anglo-Saxon authors.
[21] I also trim the sample to the papers published between 1988 and 2004.
[22] Organization fixed effects are not controlled for here but the results are consistent.
[23] The drop in last author position may due to English proficiency (Hunt, 2013) or other reasons.

Saxon first authors have higher impact factor and more citations, and non-Anglo-Saxon authors have been written higher quality paper over time. The question is, however, whether they write more/less papers than Anglo-Saxon authors do. It is possible that non-Anglo-Saxon authors only write higher quality papers at the cost of less quantity. Is it true?

The existing difficulty may be the name disambiguation problem mentioned above. First, to test whether the non-Anglo-Saxon names are more likely to be replicated, I calculate the total number of papers written by a particular author identifier (ID) and the number of years the ID appearing for all author IDs in the sample, and then conduct the following regression:

$$N_i^{appear} = \alpha_0 + \alpha_1 Eth_f + \alpha_2 Eth_l + \delta Z_i + \epsilon_i \qquad (2)$$

in which N_i^{appear} denotes the time of the author ID appearing in the 23 years and Z_i denotes a set of covariates, including average number of addresses, probability distribution of author position, publication year, states and subfields. The coefficients, α_1 and α_2, are reported in Table A5. In the first column, $Eth_{f/l}$ is the probability of the first/last author is non-Anglo-Saxon, and the coefficients are negative and significant for both first and last author position, indicating that the non-Anglo-Saxon names are less likely to be replicated or the non-Anglo-Saxon authors write fewer papers. However, the following analysis shows that the latter explanation is not true.

To test whether the non-Anglo-Saxon authors write more papers than the Anglo-Saxon authors do. I estimate the following equation:

$$N_i^{paper} = \alpha_0 + \alpha_1 Eth_f + \alpha_2 Eth_l + \delta Z_i + \gamma N_i^{appear} + \epsilon_i \qquad (3)$$

in which N_i^{paper} denotes the total number of papers written by the author with the identifier *i*. I control for the extent of replication of the author identifiers by adding the number of appearance N_i^{appear} here.[24] Results are reported in Panel A of Table 5, which shows the total number of papers written by one particular non-Anglo-Saxon author is significantly larger than that of Anglo-Saxon one. Therefore, combining the estimation results from Equation (2) and Equation (3), I conclude that the non-Anglo-Saxon authors have less likelihood of name replication and write more papers in total. I also restrict to the sample to non-popular names and report the results in the second column, which also shows a consistent pattern. To test the robustness, I calculate the total number of papers where the researcher is first or last author, and conduct the same regressions in columns 3 and 4. The results are also consistent.

[24] The results are consistent when this variable is not added.

I then calculate the number of papers written by the same author ID in each active year (at least one papers published), and conduct the following regressions:

$$N_{i,t}^{paper} = \alpha_0 + \alpha_1 Eth_f + \alpha_2 Eth_l + \delta Z_{i,t} + \gamma N_i^{appear} + \epsilon_i \qquad (4)$$

with the same settings as Equation (3). Results are reported in Panel B, and provide evidence that non-Anglo-Saxon authors write more papers than the Anglo-Saxon ones do. Considering the name replication may still exist because in there are potentially different authors with same author ID write papers in the same year and the number of years appearing may not completely capture this effect, I then further divide the number of articles written by the author (ID) by the number of years the ID appearing, use this variable as dependent variable, and report the results in Panel C. This method is likely to underestimate the productivity effect because the number of years the name appearing somehow contains information in productivity. However, the results also show that non-Anglo-Saxon authors, no matter first or last author position, write more papers.

Provided the consistent and robust results in previous analysis, I conclude here that non-Anglo-Saxon authors write more papers in total and in each active year than the Anglo-Saxon authors do.

3.4 Paper Quality of American-Born Chinese (ABC) and China-Born Chinese (CBC)

It is interesting to compare the papers produced by US-born Chinese and CHN-born Chinese. It is a test for positive selection hypothesis raised by Borjas (1986) and Gaule and Piacentini (2012) since Gaule and Piacentini (2012) argue that the positive impact from Chinese Chemistry students in US on paper quality and quantities are largely because of selection effect because the students are from the highest-level universities or colleges in China. If this is the case, then CBCs should outperform the ABCs because the latter are not selected by immigration. Comparing the outcomes of ABCs and CBCs also help to separate the effect from immigration and ethnicity provided that CBCs are mostly immigrants but ABCs are native American, though their ethnicity are both Chinese. Results from this study support the positive selection hypothesis and show that the first author premium of Chinese is contributed by immigrants.

The key hypothesis is that US-born Chinese will not be given a typical Chinese first name. Names given persons born in China are far more likely to have initials with the letters Z, Y, Q and X persons than persons born in the US, whose first names are often Anglicized. For example someone born in US might be named Richard Wang (217 people listed on white pages http://names.whitepages.com/Richard/Wang) whereas someone born in China might be named Xia Wang (58 people in white pages with this name http://names.whitepages.com/Xia/Wang). I calculate the distribution by initials of given names in US-based Chinese, US-based American and China-based Chinese, which are reported in columns 1 to 3 in Table A6. The ratios in column 4 show that letters Z, Y, Q and X are typical initials of first names. Then I assume the first names starting with these letters are CHN-born and the CHN-born researchers immigrate to US randomly in the initials of given names. Thus, I calculate a conditional probability distribution of US-born and CHN-born given the first name initials of Chinese authors. I then match this probability to the sample used in Section 2. Figure 2 plots the proportions of ABCs and CBCs by author position. Consistent with expectation, the majority of Chinese in the sample is composed of China-born Chinese (73 percent), and the Chinese proportion rising is also contributed by the CBCs.

I then conduct the same regressions with replication of Chinese by US-born Chinese and CHN-born Chinese, and present the results in Table 6. For the first author, we find the positive associated only exists significantly in CHN-born Chinese. For the last author, we find that the negative coefficients in impact factor is smaller in magnitude for China-born Chinese, and that the association between China-born Chinese and citation rank is significantly positive. In sum, the estimates in Table 6 are consistent with positive selection hypothesis, and it the first author premium of Chinese is mainly contributed by immigrants.

III. Conclusions and Discussions

The globalization of science has changed the ethnic and national origin of scientists and engineers in the US. Over half of post-doctorate workers in science labs come from overseas. The names of the authors show the Non-Anglo-Saxon proportion of first authors increases by 42 percent from 39 percent to 56 percent from 1985 to 2007, and that of last authors increases from 32 to 44 percent.

Comparing the quality measures of papers written by non-Anglo-Saxon authors and those by Anglo-Saxon ones indicates that first author being non-Anglo-Saxon is associated with significantly higher impact factor and more citations, with however the first author premium explained largely by organization fixed effects and co-authorship. The results also show that last authors being non-Anglo-Saxon is negatively associated with lower impact factor and few citations. By testing whether Anglo-Saxon first authors are relatively more highly selected to become last authors (Selection effect) and whether the Anglo-Saxon last authors have more experience in academic research (Cohort effect), I find evidence for both of them. Due to data limitation, however, I cannot specify how much either of the effects can account for.

Furthermore, dividing the papers by publication years and doing analysis in each subsample indicate that the papers written by non-Anglo-Saxon authors have been having higher quality since late 1980s. During late 1980s, Anglo-Saxon authors outperform their counterpart measured by journal impact factor and number of citations. After 2000, however, the result is opposite. Further analysis reveals the main driving force is the entry of newcomers with higher ability or talent, indicating that US has been actually attracting high-quality researchers since late 1980s.

Going beyond quality, combining the results that non-Anglo-Saxon identifiers are less likely to repeat in different years and that non-Anglo-Saxon identifiers are associated with more papers overall indicates that non-Anglo-Saxon authors write more papers in total during the period, and that they write more papers in a year when they are active (write at least one paper that year). These results suggest that non-Anglo-Saxon authors contribute a lot to the development of science research in US.

Differentiating the Chinese into American born and China born by the first name initials and doing the same analysis as above, I finds that the first author premium is mainly contributed by the China born Chinese and that Chine born Chinese outperform the American born, measured by paper quality, which is consistent with positive selection theory in previous literature. This result also implies that immigrants, the main component of non-Anglo-Saxon authors, plays an important role in scientific research in US and are more likely to do better than native ones here.

References

Borjas, George, “Self-selection and the earnings of immigrants,” *American Economic Review,* 77(1987), 531- 553

Borjas, George, Kirk B. Doran. “The Collapse of the Soviet Union and the Productivity of American Mathematicians” *The Quarterly Journal of Economics*, 2012.

Bound, John., Sarah Turner, Patrick Walsh, “Internationalization of U.S. Doctorate Education” in Science and Engineering Careers in the United States: An Analysis of Markets and Employment (2009)

Freeman, Richard B., Ina Ganguli, and Raviv Goroff-Muriciano. “International Collaboration in Research. ”, Working Paper, 2012

Freeman, Richard B., and Wei Huang. “Collaborating with People Like Me: Ethnic Coauthorship within the US”, Working Paper, 2012

Gorgger, Jeffery and Gordon Hansen. “Income maximization and the selection and sorting of international migrants.”, *Journal of Development Economics,* 2011

Gorgger, Jeffery and Gordon Hansen. “Attracting Talent: Location Choices of Foreign-Born PhDs in the US.”, NBER Working Paper, No. 18780

Hunt, Jennifer, “Which Immigrants are Most innovative and Entrepreneurial? Distinctions by Entry Visa,” NBER Working Paper, No. 14920

Jasso, Guillermina, and Mark R. Rosenzweig, “Self-Selection and the Earnings of Immigrants: Comment,” *American Economic Review*, 80(1990), 293-304.

Kerr, William R. and William F. Lincoln, “The Supply Side of Innovation: H-1B Visa Reforms and US Ethnic Invention,” *The Journal of Labor Economics* 28:3 (July 2010), 473-508.

Kerr, William R. , “Ethnic Scientific Communities and International Technology Diffusion,” *The Review of Economics and Statistics,* 90:3 (August 2008), 518-537.

NSF, “Why did they come to the United States? A profile of immigrant scientists and engineers,” National Science Foundation, Division of Science Resources Statistics, 2007

NSF, “Science and Engineering Doctorate Awards: 2006” National Science Foundation, Division of Science Resources Statistics, 2009

Stephan, Paula E. “Tabulations from Survey of Earned Doctorates,” Discussions at NBER Summer Institute (2010)

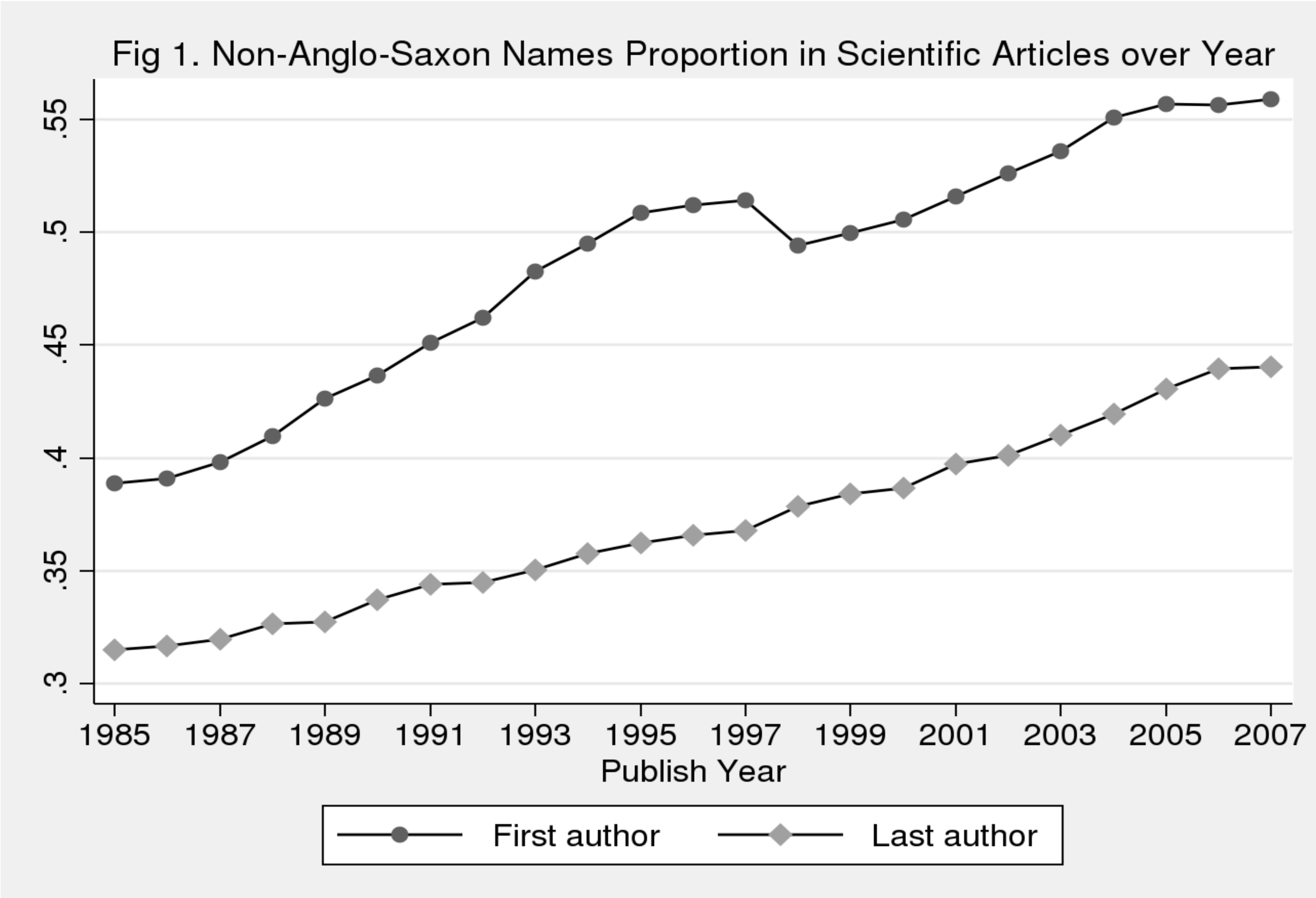

Fig 1. Non-Anglo-Saxon Names Proportion in Scientific Articles over Year
.3
.35
.4
.45
.5
.55
1985
1987
1989
1991
1993
1995
1997
1999
2001
2003
2005
2007
Publish Year
First author
Last author

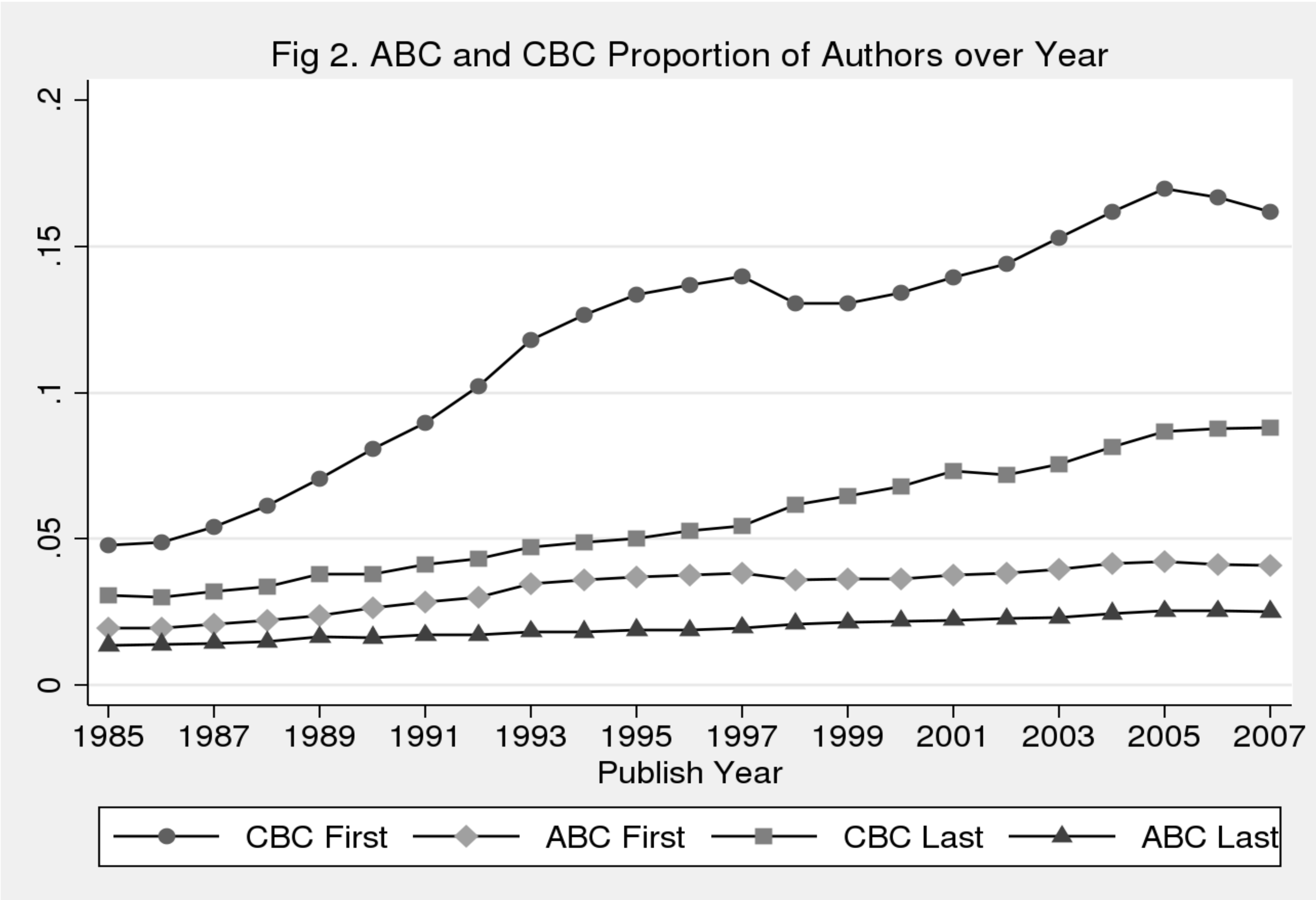
Fig 2. ABC and CBC Proportion of Authors over Year
0
.05
.1
.15
.2
1985
1987
1989
1991
1993
1995
1997
1999
2001
2003
2005
2007
Publish Year
CBC First
ABC First
CBC Last
ABC Last

Fig A1a. Non-Anglo-Saxon Names Proportion of First Authors over Year, by Ethnicities

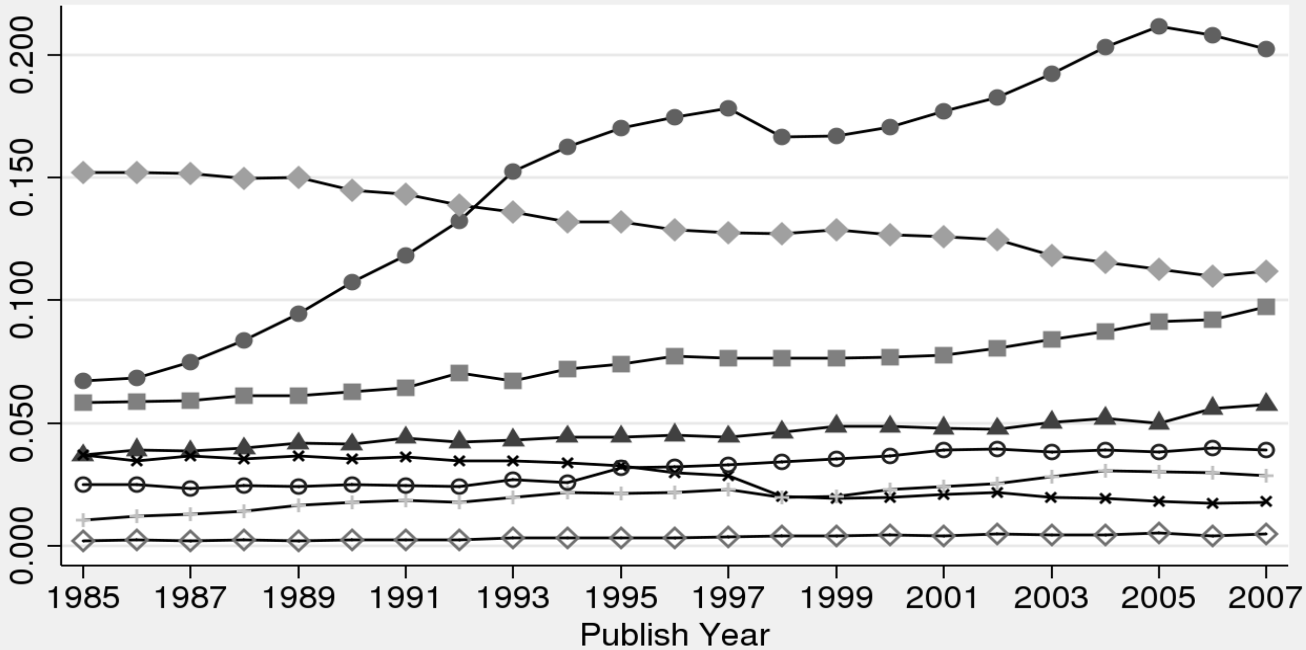

Fig A1b. Non-Anglo-Saxon Names Proportion of Last Authors over Year, by Ethnicities

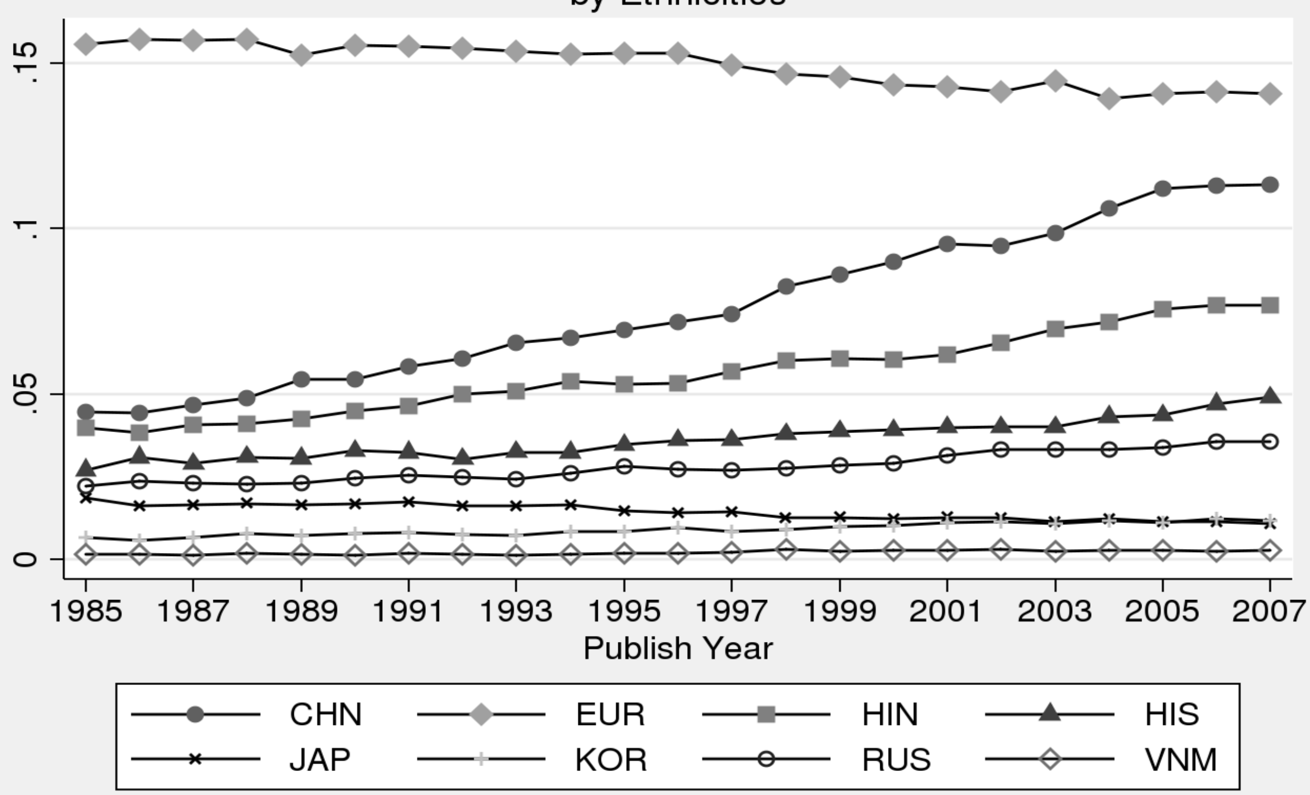

Table 1: Impact factor, Citaions and non-Anglo-Saxon authors

	(1)	(2)	(3)	(4)	(5)
Panel A: Dependent variable is impact factor					
First author non Anglo-Saxon name	0.074***	0.047***	0.037***	0.027**	
	(0.005)	(0.006)	(0.007)	(0.013)	
Last author non Anglo-Saxon name	-0.051***	-0.054***			-0.042***
	(0.005)	(0.006)			(0.008)
Observations	1,164,919	1,164,919	1,164,919	1,164,919	1,164,919
R-squared	0.392	0.541	0.717	0.886	0.791
Panel B: Dependent variable is citation rank					
First author non Anglo-Saxon name	0.784***	0.314***	0.021	-0.066	
	(0.054)	(0.068)	(0.077)	(0.155)	
Last author non Anglo-Saxon name	-0.080	-0.159**			-0.182*
	(0.055)	(0.069)			(0.104)
Observations	1,164,919	1,164,919	1,164,919	1,164,919	1,164,919
R-squared	0.175	0.394	0.593	0.824	0.677
Covariates controlled for					
Organization fixed effect	No	Yes	No	Yes	No
Last author name fixed effect	No	No	Yes	Yes	No
Interations between org and last author	No	No	No	Yes	No
First author name fixed effect	No	No	No	No	Yes
Num. of addresses	Yes	Yes	Yes	Yes	Yes
Publish year dummies	Yes	Yes	Yes	Yes	Yes
Num. of authors dummies	Yes	Yes	Yes	Yes	Yes
State dummies	Yes	Yes	Yes	Yes	Yes
Subfield dummies	Yes	Yes	Yes	Yes	Yes

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Table 2: Impact factor, Citaions and non-Anglo-Saxon authors in 1988 - 2004, by first author or last author matched or not

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Sample	Full sample	Subsamples by last author matched or not		Subsamples by first author matched or not		Full sample	Subsamples by last author matched or not		Subsamples by first author matched or not	
		Matched	Not matched	Matched	Not matched		Matched	Not matched	Matched	Not matched
Panel A: Dependent variable is impact factor										
First author non Anglo-Saxon name	0.083***	0.062***	0.089***	0.064***	0.098***	0.054***	0.044***	0.058***	0.053***	0.061***
	(0.006)	(0.008)	(0.008)	(0.009)	(0.007)	(0.007)	(0.010)	(0.011)	(0.012)	(0.009)
Last author non Anglo-Saxon name	-0.061***	-0.059***	-0.067***	-0.061***	-0.065***	-0.064***	-0.064***	-0.066***	-0.051***	-0.073***
	(0.006)	(0.008)	(0.008)	(0.009)	(0.007)	(0.007)	(0.010)	(0.011)	(0.012)	(0.010)
Observations	853,840	398,058	455,782	326,118	527,722	853,840	398,058	455,782	326,118	527,722
R-squared	0.408	0.380	0.431	0.389	0.421	0.553	0.553	0.580	0.562	0.571
Panel B: Dependent variable is citation rank										
First author non Anglo-Saxon name	0.708***	0.464***	0.657***	-0.002	1.355***	0.236***	0.264**	0.071	-0.073	0.722***
	(0.061)	(0.090)	(0.084)	(0.101)	(0.077)	(0.078)	(0.117)	(0.111)	(0.135)	(0.100)
Last author non Anglo-Saxon name	-0.248***	-0.261***	-0.396***	-0.532***	-0.289***	-0.343***	-0.175	-0.619***	-0.517***	-0.414***
	(0.063)	(0.092)	(0.086)	(0.102)	(0.080)	(0.080)	(0.121)	(0.114)	(0.137)	(0.103)
Observations	853,840	398,058	455,782	326,118	527,722	853,840	398,058	455,782	326,118	527,722
R-squared	0.171	0.155	0.187	0.154	0.190	0.396	0.400	0.430	0.416	0.421
Covariates controlled for										
Organization fixed effect	No	No	No	No	No	Yes	Yes	Yes	Yes	Yes
Num. of addresses	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Publish year dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Num. of authors dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
State dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Subfield dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Table 3: Impact factor, Citaions and non-Anglo-Saxon authors, by publish years

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Publish year	1985-1990	1991-1995	1996-2000	2001-2007	1985-1990	1991-1995	1996-2000	2001-2007
Panel A: Dependent variable is impact factor								
First author non Anglo-Saxon name	-0.003	0.065***	0.090***	0.094***	-0.010	0.035**	0.059***	0.065***
	(0.009)	(0.011)	(0.009)	(0.007)	(0.011)	(0.014)	(0.013)	(0.010)
Last author non Anglo-Saxon name	-0.050***	-0.093***	-0.069***	-0.027***	-0.046***	-0.098***	-0.079***	-0.026**
	(0.009)	(0.011)	(0.010)	(0.007)	(0.012)	(0.014)	(0.013)	(0.010)
Observations	214,529	208,036	330,522	455,160	214,529	208,036	330,522	455,160
R-squared	0.422	0.440	0.397	0.386	0.546	0.569	0.564	0.573
Panel B: Dependent variable is citation rank								
First author non Anglo-Saxon name	-0.776***	-0.060	0.786***	1.735***	-0.947***	-0.499***	0.396***	1.214***
	(0.124)	(0.123)	(0.098)	(0.088)	(0.150)	(0.153)	(0.130)	(0.121)
Last author non Anglo-Saxon name	-0.917***	-1.158***	-0.284***	0.664***	-0.877***	-1.228***	-0.422***	0.690***
	(0.130)	(0.128)	(0.101)	(0.088)	(0.158)	(0.160)	(0.134)	(0.120)
Observations	214,529	208,036	330,522	455,160	214,529	208,036	330,522	455,160
R-squared	0.147	0.165	0.176	0.205	0.338	0.368	0.433	0.462
Covariates controlled for								
Organization fixed effect	No	No	No	No	Yes	Yes	Yes	Yes
Num. of addresses	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Publish year dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Num. of authors dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
State dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Subfield dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Table 4: Impact factor, citation rank and non-Anglo-Saxon authors, by publish year and first or last author appearing first time

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Dependent variable is impact factor				Dependent variable is citation rank			
Sample	First author new name	First author not new name	Last author new name	Last author not new name	First author new name	First author not new name	Last author new name	Last author not new name
Panel A. 1988 - 1990 papers								
First author non-Anglo-Saxon	0.006	-0.007	-0.006	-0.031*	-0.525**	-1.134***	-0.819***	-1.321***
	(0.017)	(0.020)	(0.020)	(0.017)	(0.213)	(0.284)	(0.254)	(0.229)
Last author non-Anglo-Saxon	-0.069***	-0.048**	-0.029	-0.068***	-0.577**	-1.900***	-0.414	-1.309***
	(0.018)	(0.021)	(0.020)	(0.018)	(0.225)	(0.294)	(0.262)	(0.244)
Observations	71,606	40,542	52,940	59,208	71,606	40,542	52,940	59,208
R-squared	0.483	0.418	0.461	0.471	0.164	0.137	0.153	0.145
Panel B.1991-1995 papers								
First author non-Anglo-Saxon	0.072***	0.056***	0.067***	0.023	0.364**	-0.588***	-0.095	-0.582***
	(0.015)	(0.016)	(0.018)	(0.014)	(0.164)	(0.187)	(0.203)	(0.154)
Last author non-Anglo-Saxon	-0.100***	-0.085***	-0.066***	-0.083***	-1.162***	-1.258***	-0.570***	-1.241***
	(0.016)	(0.016)	(0.018)	(0.015)	(0.172)	(0.192)	(0.206)	(0.163)
Observations	116,718	91,318	79,139	128,897	116,718	91,318	79,139	128,897
R-squared	0.468	0.395	0.444	0.440	0.185	0.143	0.168	0.155
Panel C. 1996-2000 papers								
First author non-Anglo-Saxon	0.100***	0.075***	0.080***	0.047***	1.334***	0.361**	0.788***	0.274**
	(0.015)	(0.013)	(0.017)	(0.013)	(0.148)	(0.149)	(0.177)	(0.130)
Last author non-Anglo-Saxon	-0.070***	-0.061***	-0.001	-0.063***	-0.141	-0.366**	0.720***	-0.221
	(0.016)	(0.014)	(0.017)	(0.013)	(0.154)	(0.151)	(0.179)	(0.134)
Observations	143,081	144,113	103,113	184,081	143,081	144,113	103,113	184,081
R-squared	0.413	0.374	0.371	0.409	0.201	0.158	0.173	0.171
Panel D. 2001-2004 papers								
First author non-Anglo-Saxon	0.127***	0.106***	0.131***	0.068***	1.995***	1.451***	1.527***	1.221***
	(0.016)	(0.012)	(0.018)	(0.012)	(0.173)	(0.154)	(0.204)	(0.138)
Last author non-Anglo-Saxon	-0.026	-0.045***	0.003	-0.025**	0.519***	0.503***	1.299***	0.586***
	(0.017)	(0.012)	(0.018)	(0.012)	(0.176)	(0.153)	(0.204)	(0.139)
Observations	106,716	139,746	78,494	167,968	106,716	139,746	78,494	167,968
R-squared	0.419	0.389	0.377	0.421	0.226	0.184	0.194	0.196
Covariates controlled for								
Organization fixed effect	No	No	No	No	No	No	No	No
Num. of addresses	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Publish year dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Num. of authors dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
State dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Subfield dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Table 5: Number of papers written and non-Anglo-Saxon authors

	(1)	(2)	(3)	(4)
	All papers calculated		Only first authored or last authored paper calculated	
	Full sample	Time of Author ID Appearing <= 5	Full sample	Time of Author ID Appearing <= 5
Panel A: Dependent variable is Total number of papers written				
First author non Anglo-Saxon name	0.132***	0.047***	0.093***	0.045***
	(0.005)	(0.002)	(0.004)	(0.002)
Last author non Anglo-Saxon name	0.177***	0.032***	0.164***	0.033***
	(0.010)	(0.003)	(0.010)	(0.003)
Observations	1,110,223	1,017,559	1,110,223	1,017,559
R-squared	0.863	0.832	0.824	0.776
Panel B: Dependent variable is Number of papers written in a particular year conditional active				
First author non Anglo-Saxon name	0.053***	0.026***	0.043***	0.025***
	(0.001)	(0.001)	(0.001)	(0.001)
Last author non Anglo-Saxon name	0.059***	0.018***	0.053***	0.018***
	(0.002)	(0.001)	(0.002)	(0.001)
Observations	2,533,030	1,675,885	2,533,030	1,675,885
R-squared	0.193	0.039	0.506	0.704
Panel C: Dependent variable is Number of papers written in a particular year conditional on active/Time of Name ID appearing				
First author non Anglo-Saxon name	0.011***	0.011***	0.011***	0.014***
	(0.000)	(0.001)	(0.000)	(0.000)
Last author non Anglo-Saxon name	0.007***	0.006***	0.008***	0.011***
	(0.000)	(0.001)	(0.000)	(0.000)
Observations	2,533,030	1,675,885	857,145	2,533,030
R-squared	0.792	0.704	0.110	0.676
Covariates controlled for				
Appearing time dummies	Yes	Yes	Yes	Yes
Num. of addresses	Yes	Yes	Yes	Yes
Publish year dummies	Yes	Yes	Yes	Yes
Num. of authors dummies	Yes	Yes	Yes	Yes
State dummies	Yes	Yes	Yes	Yes
Subfield dummies	Yes	Yes	Yes	Yes

Table 6: Impact factor, Citaions and Chinese Names

	(1)	(2)	(3)	(4)
	Dependent variable is impact factor		Dependent variable is citation rank	
Panel A: Dpeendent variable is impact factor				
First author China Born Chinese	0.253***	0.212***	3.074***	2.190***
	(0.010)	(0.013)	(0.113)	(0.140)
First author American Born Chinese	0.010	-0.035	-1.390***	-1.756***
	(0.027)	(0.033)	(0.306)	(0.374)
Last author China Born Chinese	-0.071***	-0.057***	1.357***	1.470***
	(0.013)	(0.017)	(0.168)	(0.209)
Last author American Born Chinese	-0.205***	-0.211***	-3.180***	-3.512***
	(0.032)	(0.041)	(0.410)	(0.509)
Observations	1,164,919	1,164,919	1,164,919	1,164,919
R-squared	0.393	0.542	0.176	0.395
Covariates controlled for				
Organization fixed effect	No	Yes	No	Yes
Other ethnicities for first and last authors	Yes	Yes	Yes	Yes
Num. of addresses	Yes	Yes	Yes	Yes
Publish year dummies	Yes	Yes	Yes	Yes
Num. of authors dummies	Yes	Yes	Yes	Yes
State dummies	Yes	Yes	Yes	Yes
Subfield dummies	Yes	Yes	Yes	Yes

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Table A1: Summary statistics

Variable	(1) Full sample	(2) Subsamples by number of authors: Two-author	(3) Three-author	(4) Four-author
Panel A: Author ethnicity information				
First author Non-Anglo-Saxon name	0.50	0.50	0.50	0.49
	(0.50)	(0.50)	(0.50)	(0.50)
Last author Non-Anglo-Saxon name	0.38	0.39	0.37	0.36
	(0.48)	(0.49)	(0.48)	(0.48)
Panel B: Article information				
Two-year impact factor	2.57	2.31	2.60	2.93
	(3.03)	(2.89)	(3.00)	(3.25)
Citation rank	47.62	44.21	48.21	52.20
	(30.50)	(30.52)	(30.30)	(30.08)
Number of references	29.41	29.29	29.26	29.82
	(18.96)	(19.91)	(18.51)	(17.98)
Number of addresses	1.63	1.37	1.67	1.97
	(0.87)	(0.62)	(0.86)	(1.07)
Panel C: Information of authors' previous articles				
Average impact factor of previous papers				
First author	2.56	2.41	2.58	2.75
	(1.89)	(1.84)	(1.89)	(1.95)
Last author	2.76	2.55	2.79	3.03
	(2.20)	(2.15)	(2.18)	(2.27)
Number of first author's previous papers				
None	0.35	0.36	0.35	0.33
	(0.48)	(0.48)	(0.48)	(0.47)
One to ten	0.49	0.48	0.49	0.50
	(0.50)	(0.50)	(0.50)	(0.50)
Eleven and above	0.17	0.16	0.16	0.17
	(0.37)	(0.37)	(0.37)	(0.38)
Number of last author's previous papers				
None	0.20	0.20	0.20	0.19
	(0.40)	(0.40)	(0.40)	(0.39)
One to ten	0.42	0.43	0.41	0.40
	(0.49)	(0.49)	(0.49)	(0.49)
Eleven and above	0.38	0.37	0.38	0.40
	(0.49)	(0.48)	(0.49)	(0.49)
Observations	1,164,919	467,882	400,800	296,237

Note: Data source is Web of Science. Standard deviations in parentheses

Table A2: Impact factor, Citaions and non-Anglo-Saxon authors

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Dependent variable	Impact factor					Citations rank				
First author ethnicities										
CHN	0.200***	0.159***	0.109***	0.079***		2.111***	1.342***	0.604***	0.287	
	(0.007)	(0.009)	(0.009)	(0.018)		(0.079)	(0.098)	(0.110)	(0.214)	
EUR	0.032***	0.007	-0.012	-0.019		0.625***	0.329***	0.003	-0.112	
	(0.007)	(0.009)	(0.010)	(0.020)		(0.082)	(0.104)	(0.116)	(0.238)	
HIN	-0.074***	-0.060***	-0.004	0.002		-1.315***	-1.266***	-0.748***	-0.589**	
	(0.008)	(0.011)	(0.012)	(0.023)		(0.103)	(0.130)	(0.147)	(0.295)	
HIS	-0.004	-0.036***	-0.005	-0.004		-0.226*	-0.601***	-0.748***	-0.752**	
	(0.011)	(0.014)	(0.015)	(0.029)		(0.129)	(0.161)	(0.178)	(0.355)	
JAP	0.294***	0.174***	0.116***	0.065		2.925***	1.562***	0.447**	0.388	
	(0.017)	(0.020)	(0.022)	(0.039)		(0.165)	(0.199)	(0.221)	(0.415)	
KOR	0.121***	0.070***	0.066***	0.029		0.876***	-0.074	0.035	0.032	
	(0.015)	(0.019)	(0.019)	(0.034)		(0.185)	(0.224)	(0.244)	(0.454)	
RUS	0.026**	-0.003	-0.003	0.017		0.570***	-0.055	0.101	0.008	
	(0.013)	(0.017)	(0.018)	(0.036)		(0.154)	(0.193)	(0.216)	(0.435)	
VNM	0.044	0.056	0.031	0.059		-0.843*	-0.588	-0.775	0.038	
	(0.042)	(0.053)	(0.053)	(0.087)		(0.457)	(0.562)	(0.601)	(1.145)	
Last author ethnicities										
CHN	-0.104***	-0.095***			-0.059***	0.228**	0.226*			0.071
	(0.008)	(0.010)			(0.014)	(0.102)	(0.128)			(0.187)
EUR	0.027***	0.010			0.006	0.374***	0.161*			0.199
	(0.007)	(0.009)			(0.012)	(0.076)	(0.096)			(0.144)
HIN	-0.192***	-0.153***			-0.126***	-1.624***	-1.379***			-1.160***
	(0.009)	(0.011)			(0.015)	(0.116)	(0.146)			(0.221)
HIS	-0.092***	-0.089***			-0.068***	-0.409***	-0.128			-0.632**
	(0.012)	(0.015)			(0.020)	(0.144)	(0.182)			(0.268)
JAP	-0.101***	-0.143***			-0.078**	-1.211***	-1.688***			-1.408***
	(0.020)	(0.025)			(0.037)	(0.227)	(0.279)			(0.433)
KOR	-0.186***	-0.189***			-0.141***	-0.973***	-1.534***			-0.505
	(0.022)	(0.028)			(0.035)	(0.286)	(0.353)			(0.503)
RUS	-0.008	-0.018			-0.003	0.458***	0.284			0.280
	(0.015)	(0.019)			(0.026)	(0.164)	(0.208)			(0.313)
VNM	-0.160***	-0.117**			-0.065	-1.479**	-0.658			-1.377
	(0.045)	(0.057)			(0.079)	(0.587)	(0.731)			(1.027)
Observations	1,164,919	1,164,919	1,164,919	1,164,919	1,164,919	1,164,919	1,164,919	1,164,919	1,164,919	1,164,919
R-squared	0.393	0.542	0.717	0.886	0.791	0.176	0.395	0.593	0.824	0.678
Covariates controlled for										
Organization fixed effect	No	Yes	No	Yes	No	No	Yes	No	Yes	No
Last author name fixed effect	No	No	Yes	Yes	No	No	No	Yes	Yes	No
Interations between org and last author	No	No	No	Yes	No	No	No	No	Yes	No
First author name fixed effect	No	No	No	No	Yes	No	No	No	No	Yes
Num. of addresses	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Publish year dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Num. of authors dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
State dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Subfield dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Table A3a: Impact factor, citation and non-Anglo-Saxon authors, by fields

	(1) Multidisciplinary	(2) Agriculture	(3) Biology	(4) Biomedicine	(5) Chemistry	(6) Clinical Medicine	(7) Engineering	(8) Geosciences	(9) ICT	(10) Material	(11) Mathmatics	(12) Physics
Panel A: Dependent variable is impact factor												
First author non Anglo-Saxon name	-0.057	0.063***	0.091***	0.063***	-0.020**	0.078***	0.014**	-0.016	0.001	-0.040**	-0.006	0.023**
	(0.142)	(0.007)	(0.022)	(0.014)	(0.010)	(0.011)	(0.006)	(0.011)	(0.006)	(0.019)	(0.008)	(0.010)
Last author non Anglo-Saxon name	-0.629***	0.009	0.017	-0.063***	-0.061***	-0.014	-0.009*	-0.018	-0.004	-0.007	-0.000	0.010
	(0.146)	(0.008)	(0.024)	(0.015)	(0.010)	(0.011)	(0.006)	(0.012)	(0.006)	(0.018)	(0.008)	(0.010)
Observations	23,342	112,641	75,418	297,519	128,728	340,491	98,397	47,369	89,199	42,011	36,629	134,829
R-squared	0.510	0.639	0.672	0.438	0.568	0.512	0.681	0.644	0.674	0.599	0.819	0.576
Panel B: Dependent variable is citation rank												
First author non Anglo-Saxon name	1.404***	0.916***	-0.352	0.352***	0.719***	0.111	1.247***	-1.587***	0.030	0.823**	0.088	0.320
	(0.533)	(0.224)	(0.292)	(0.128)	(0.189)	(0.138)	(0.262)	(0.395)	(0.285)	(0.410)	(0.478)	(0.214)
Last author non Anglo-Saxon name	-0.444	0.141	0.218	-0.419***	0.269	0.043	1.036***	-0.934**	0.730***	0.936**	0.423	0.223
	(0.551)	(0.244)	(0.318)	(0.132)	(0.195)	(0.143)	(0.257)	(0.419)	(0.273)	(0.397)	(0.477)	(0.211)
Observations	23,342	112,641	75,418	297,519	128,728	340,491	98,397	47,369	89,199	42,011	36,629	134,829
R-squared	0.531	0.409	0.498	0.358	0.345	0.403	0.466	0.485	0.440	0.443	0.494	0.368
Covariates controlled for												
Organization fixed effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Num. of addresses	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Publish year dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Num. of authors dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
State dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Subfield dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Robust standard errors in parentheses

Table A3b: Proportion of immigrants, by fields

	(1) Multidisciplinary	(2) Agriculture	(3) Biology	(4) Biomedicine	(5) Chemistry	(6) Clinical Medicine	(7) Engineering	(8) Geosciences	(9) ICT	(10) Material	(11) Mathmatics	(12) Physics
Panel A: First author												
Non Anglo-Saxon	0.50	0.39	0.35	0.52	0.59	0.43	0.59	0.41	0.63	0.64	0.54	0.58
	(0.50)	(0.49)	(0.47)	(0.50)	(0.49)	(0.49)	(0.49)	(0.49)	(0.48)	(0.48)	(0.49)	(0.49)
By Ethnicity												
CHN	0.15	0.11	0.08	0.18	0.25	0.09	0.23	0.13	0.26	0.28	0.20	0.22
	(0.36)	(0.31)	(0.27)	(0.38)	(0.43)	(0.29)	(0.41)	(0.33)	(0.43)	(0.44)	(0.39)	(0.41)
EUR	0.14	0.13	0.13	0.13	0.11	0.15	0.11	0.13	0.10	0.09	0.13	0.13
	(0.34)	(0.32)	(0.33)	(0.32)	(0.31)	(0.34)	(0.30)	(0.33)	(0.30)	(0.28)	(0.33)	(0.33)
HIN	0.06	0.05	0.03	0.07	0.09	0.07	0.12	0.04	0.13	0.12	0.08	0.08
	(0.23)	(0.22)	(0.18)	(0.26)	(0.28)	(0.25)	(0.33)	(0.19)	(0.34)	(0.33)	(0.27)	(0.27)
HIS	0.05	0.05	0.05	0.05	0.04	0.05	0.04	0.04	0.04	0.04	0.05	0.04
	(0.20)	(0.22)	(0.22)	(0.21)	(0.19)	(0.21)	(0.20)	(0.19)	(0.19)	(0.18)	(0.20)	(0.20)
JAP	0.04	0.01	0.02	0.04	0.02	0.03	0.02	0.02	0.01	0.02	0.01	0.02
	(0.20)	(0.12)	(0.13)	(0.19)	(0.15)	(0.16)	(0.13)	(0.13)	(0.12)	(0.15)	(0.10)	(0.15)
KOR	0.02	0.02	0.01	0.02	0.03	0.01	0.04	0.02	0.04	0.05	0.02	0.03
	(0.14)	(0.14)	(0.10)	(0.13)	(0.17)	(0.11)	(0.19)	(0.12)	(0.17)	(0.20)	(0.13)	(0.17)
RUS	0.04	0.02	0.02	0.03	0.04	0.03	0.03	0.03	0.04	0.04	0.05	0.06
	(0.18)	(0.13)	(0.14)	(0.17)	(0.18)	(0.16)	(0.16)	(0.16)	(0.18)	(0.18)	(0.22)	(0.22)
VNM	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00
	(0.05)	(0.04)	(0.05)	(0.06)	(0.07)	(0.06)	(0.06)	(0.04)	(0.07)	(0.06)	(0.06)	(0.05)
Panel B: Last author												
Non Anglo-Saxon	0.36	0.28	0.27	0.37	0.40	0.35	0.47	0.32	0.54	0.47	0.51	0.46
	(0.48)	(0.45)	(0.44)	(0.48)	(0.49)	(0.47)	(0.50)	(0.46)	(0.50)	(0.50)	(0.50)	(0.49)
By Ethnicity												
CHN	0.07	0.05	0.03	0.07	0.10	0.05	0.13	0.08	0.17	0.14	0.18	0.13
	(0.24)	(0.21)	(0.18)	(0.25)	(0.29)	(0.21)	(0.33)	(0.26)	(0.37)	(0.34)	(0.38)	(0.33)
EUR	0.16	0.13	0.14	0.16	0.15	0.16	0.13	0.13	0.13	0.12	0.14	0.15
	(0.36)	(0.33)	(0.34)	(0.35)	(0.35)	(0.36)	(0.33)	(0.33)	(0.32)	(0.32)	(0.34)	(0.35)
HIN	0.04	0.04	0.02	0.05	0.05	0.05	0.11	0.03	0.14	0.11	0.08	0.07
	(0.19)	(0.18)	(0.15)	(0.21)	(0.22)	(0.21)	(0.31)	(0.18)	(0.34)	(0.31)	(0.26)	(0.26)
HIS	0.04	0.03	0.04	0.04	0.03	0.04	0.04	0.03	0.04	0.03	0.04	0.04
	(0.18)	(0.17)	(0.18)	(0.18)	(0.18)	(0.19)	(0.19)	(0.17)	(0.18)	(0.17)	(0.19)	(0.18)
JAP	0.02	0.01	0.01	0.02	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.02
	(0.13)	(0.10)	(0.10)	(0.13)	(0.12)	(0.11)	(0.11)	(0.10)	(0.12)	(0.11)	(0.10)	(0.12)
KOR	0.01	0.01	0.00	0.01	0.01	0.01	0.02	0.01	0.02	0.02	0.01	0.01
	(0.09)	(0.08)	(0.06)	(0.08)	(0.10)	(0.08)	(0.13)	(0.08)	(0.12)	(0.12)	(0.10)	(0.10)
RUS	0.03	0.02	0.02	0.03	0.03	0.03	0.03	0.02	0.03	0.03	0.05	0.04
	(0.17)	(0.12)	(0.13)	(0.15)	(0.17)	(0.15)	(0.16)	(0.15)	(0.18)	(0.17)	(0.21)	(0.19)
VNM	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	(0.03)	(0.04)	(0.03)	(0.04)	(0.05)	(0.04)	(0.05)	(0.03)	(0.06)	(0.05)	(0.06)	(0.05)

Table A4: Impact factor, Citaions and non-Anglo-Saxon authors of different ethnicities in 1988 - 2004, by first author or last author matched or not

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Sample	Full sample	Subsamples by last author matched or not		Subsamples by first author matched or not		Full sample	Subsamples by last author matched or not		Subsamples by first author matched or not	
		Matched	Not matched	Matched	Not matched		Matched	Not matched	Matched	Not matched
Panel A: Dependent variable is impact factor										
First author										
CHN	0.219***	0.162***	0.256***	0.165***	0.240***	0.176***	0.139***	0.204***	0.151***	0.182***
	(0.008)	(0.011)	(0.012)	(0.013)	(0.011)	(0.010)	(0.014)	(0.016)	(0.017)	(0.014)
EUR	0.037***	0.023*	0.045***	0.046***	0.043***	0.011	-0.010	0.021	0.021	0.015
	(0.008)	(0.012)	(0.012)	(0.014)	(0.011)	(0.011)	(0.016)	(0.016)	(0.019)	(0.014)
HIN	-0.076***	-0.060***	-0.102***	-0.078***	-0.073***	-0.061***	-0.047**	-0.075***	-0.055**	-0.058***
	(0.010)	(0.015)	(0.014)	(0.016)	(0.013)	(0.013)	(0.019)	(0.019)	(0.022)	(0.017)
HIS	0.009	-0.031	0.038**	0.016	0.027*	-0.027	-0.053**	-0.012	-0.012	-0.018
	(0.014)	(0.019)	(0.019)	(0.025)	(0.016)	(0.017)	(0.024)	(0.026)	(0.034)	(0.021)
JAP	0.295***	0.339***	0.248***	0.176***	0.362***	0.166***	0.251***	0.097***	0.101*	0.214***
	(0.020)	(0.029)	(0.027)	(0.042)	(0.023)	(0.024)	(0.036)	(0.034)	(0.055)	(0.028)
KOR	0.117***	0.053**	0.163***	0.060**	0.149***	0.064***	0.017	0.115***	0.067*	0.071**
	(0.018)	(0.023)	(0.028)	(0.030)	(0.023)	(0.022)	(0.029)	(0.036)	(0.039)	(0.029)
RUS	0.035**	0.040*	0.018	0.061**	0.036*	0.003	0.013	-0.013	0.023	0.010
	(0.016)	(0.022)	(0.024)	(0.026)	(0.020)	(0.021)	(0.029)	(0.032)	(0.037)	(0.027)
VNM	0.053	0.098	-0.009	-0.088	0.121**	0.054	0.113	-0.022	-0.090	0.118*
	(0.047)	(0.062)	(0.071)	(0.073)	(0.059)	(0.059)	(0.076)	(0.095)	(0.105)	(0.072)
Last author										
CHN	-0.125***	-0.117***	-0.170***	-0.145***	-0.119***	-0.117***	-0.096***	-0.169***	-0.132***	-0.103***
	(0.010)	(0.013)	(0.015)	(0.014)	(0.013)	(0.013)	(0.017)	(0.020)	(0.019)	(0.018)
EUR	0.023***	0.027**	0.023**	0.045***	0.007	0.004	-0.006	0.014	0.036**	-0.019
	(0.008)	(0.011)	(0.011)	(0.013)	(0.010)	(0.010)	(0.015)	(0.015)	(0.018)	(0.013)
HIN	-0.214***	-0.213***	-0.236***	-0.212***	-0.219***	-0.168***	-0.165***	-0.186***	-0.161***	-0.172***
	(0.010)	(0.014)	(0.015)	(0.016)	(0.014)	(0.014)	(0.019)	(0.021)	(0.022)	(0.018)
HIS	-0.094***	-0.026	-0.125***	-0.109***	-0.085***	-0.100***	-0.059**	-0.117***	-0.093***	-0.100***
	(0.014)	(0.021)	(0.019)	(0.021)	(0.019)	(0.018)	(0.028)	(0.026)	(0.029)	(0.025)
JAP	-0.109***	-0.144***	-0.080**	-0.070*	-0.133***	-0.156***	-0.189***	-0.134***	-0.080	-0.195***
	(0.025)	(0.035)	(0.034)	(0.040)	(0.031)	(0.031)	(0.044)	(0.045)	(0.054)	(0.039)
KOR	-0.190***	-0.262***	-0.131***	-0.225***	-0.179***	-0.192***	-0.248***	-0.125**	-0.200***	-0.183***
	(0.027)	(0.034)	(0.043)	(0.037)	(0.038)	(0.035)	(0.045)	(0.059)	(0.052)	(0.050)
RUS	-0.009	0.006	-0.003	0.003	-0.020	-0.012	0.002	-0.004	0.028	-0.034
	(0.018)	(0.026)	(0.024)	(0.027)	(0.023)	(0.024)	(0.035)	(0.034)	(0.038)	(0.032)
VNM	-0.159***	-0.206**	-0.094	-0.037	-0.247***	-0.127*	-0.159	-0.054	-0.042	-0.193**
	(0.055)	(0.097)	(0.067)	(0.099)	(0.062)	(0.070)	(0.122)	(0.091)	(0.121)	(0.090)
Observations	853,840	398,058	455,782	326,118	527,722	853,840	398,058	455,782	326,118	527,722
R-squared	0.409	0.381	0.432	0.389	0.422	0.553	0.554	0.580	0.563	0.571
Panel B: Dependent variable is citation rank										
First author										
CHN	2.110***	1.582***	2.186***	0.339**	2.963***	1.319***	1.248***	1.165***	0.094	1.950***
	(0.089)	(0.126)	(0.126)	(0.140)	(0.115)	(0.111)	(0.161)	(0.163)	(0.185)	(0.147)
EUR	0.560***	0.468***	0.599***	0.730***	0.892***	0.272**	0.225	0.282*	0.613***	0.541***
	(0.094)	(0.140)	(0.127)	(0.156)	(0.117)	(0.120)	(0.184)	(0.168)	(0.213)	(0.152)
HIN	-1.546***	-1.359***	-1.892***	-2.500***	-0.911***	-1.447***	-1.167***	-1.758***	-1.982***	-1.019***
	(0.120)	(0.173)	(0.165)	(0.200)	(0.149)	(0.151)	(0.224)	(0.217)	(0.269)	(0.192)
HIS	-0.371**	-1.040***	0.037	0.422	0.259	-0.782***	-1.197***	-0.610**	0.340	-0.327
	(0.148)	(0.212)	(0.206)	(0.280)	(0.174)	(0.186)	(0.274)	(0.269)	(0.376)	(0.223)
JAP	2.792***	2.853***	2.528***	1.170***	4.592***	1.325***	1.553***	0.935***	0.627	2.779***
	(0.187)	(0.270)	(0.257)	(0.407)	(0.211)	(0.227)	(0.338)	(0.322)	(0.520)	(0.262)
KOR	0.529**	0.356	0.264	-1.280***	1.728***	-0.425	0.069	-1.016***	-1.519***	0.412
	(0.213)	(0.299)	(0.301)	(0.354)	(0.266)	(0.260)	(0.375)	(0.380)	(0.455)	(0.331)
RUS	0.638***	0.559**	0.527**	1.400***	0.817***	0.051	0.264	-0.068	1.258***	0.193
	(0.176)	(0.251)	(0.246)	(0.315)	(0.211)	(0.222)	(0.325)	(0.322)	(0.427)	(0.270)
VNM	-0.664	0.168	-1.752**	-2.588**	0.835	-0.210	0.939	-1.612*	-1.570	1.045
	(0.526)	(0.729)	(0.750)	(1.020)	(0.611)	(0.648)	(0.926)	(0.955)	(1.378)	(0.756)
Last author										
CHN	-0.139	-0.635***	-0.509***	-0.576***	-0.447***	-0.111	-0.102	-0.845***	-0.400*	-0.371*
	(0.119)	(0.160)	(0.178)	(0.182)	(0.156)	(0.149)	(0.208)	(0.232)	(0.240)	(0.200)
EUR	0.261***	0.543***	0.154	0.208	0.259**	-0.003	0.214	-0.130	0.007	-0.050
	(0.087)	(0.130)	(0.117)	(0.142)	(0.110)	(0.111)	(0.172)	(0.155)	(0.192)	(0.143)
HIN	-1.682***	-1.995***	-2.031***	-2.152***	-1.605***	-1.392***	-1.454***	-1.848***	-1.822***	-1.296***
	(0.134)	(0.188)	(0.191)	(0.219)	(0.169)	(0.171)	(0.247)	(0.256)	(0.295)	(0.220)
HIS	-0.571***	0.255	-0.753***	-0.864***	-0.465**	-0.251	0.574*	-0.534*	-0.496	-0.194
	(0.167)	(0.258)	(0.218)	(0.274)	(0.209)	(0.213)	(0.337)	(0.291)	(0.370)	(0.271)
JAP	-1.588***	-1.983***	-1.130***	-1.146***	-1.946***	-2.115***	-2.160***	-1.976***	-1.274**	-2.621***
	(0.258)	(0.388)	(0.346)	(0.432)	(0.319)	(0.320)	(0.499)	(0.441)	(0.567)	(0.403)
KOR	-0.797**	-2.270***	0.165	-0.685	-1.599***	-1.390***	-2.538***	-0.228	-1.443**	-1.901***
	(0.330)	(0.449)	(0.485)	(0.503)	(0.432)	(0.408)	(0.574)	(0.624)	(0.655)	(0.547)
RUS	0.319*	0.968***	0.066	0.210	0.227	0.081	0.886**	-0.403	0.127	-0.066
	(0.190)	(0.284)	(0.254)	(0.311)	(0.238)	(0.242)	(0.374)	(0.337)	(0.417)	(0.310)
VNM	-1.687**	-2.831**	-0.399	-2.638**	-1.479*	-0.990	-2.053	0.241	-2.019	-0.818
	(0.671)	(1.169)	(0.816)	(1.060)	(0.859)	(0.845)	(1.475)	(1.086)	(1.425)	(1.105)
Observations	853,840	398,058	455,782	326,118	527,722	853,840	398,058	455,782	326,118	527,722
R-squared	0.172	0.156	0.188	0.155	0.191	0.397	0.401	0.431	0.416	0.422
Covariates controlled for										
Organization fixed effect	No	No	No	No	No	Yes	Yes	Yes	Yes	Yes
Num. of addresses	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Publish year dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Num. of authors dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
State dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Subfield dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Robust standard errors in parentheses

Table A5: Time of Author ID appearing and ethnicities

VARIABLES	(1)	(2)
	Time of Author ID Appearing (1-24)	
First author non Anglo-Saxon name	-0.025***	
	(0.006)	
First author		
CHN		0.545***
		(0.012)
EUR		-0.218***
		(0.008)
HIN		-0.011
		(0.011)
HIS		-0.420***
		(0.010)
JAP		-0.309***
		(0.015)
KOR		0.281***
		(0.028)
RUS		-0.256***
		(0.013)
VNM		-0.298***
		(0.040)
Last author non Anglo-Saxon name	-0.082***	
	(0.013)	
Last author		
CHN		0.577***
		(0.030)
EUR		-0.155***
		(0.018)
HIN		0.127***
		(0.029)
HIS		-0.633***
		(0.025)
JAP		-0.138***
		(0.053)
KOR		0.331***
		(0.080)
RUS		-0.384***
		(0.033)
VNM		-0.704***
		(0.097)
Observations	1,110,223	1,110,223
R-squared	0.049	0.054
Covariates controlled for		
Author position	Yes	Yes
Num. of addresses	Yes	Yes
Publish year dummies	Yes	Yes
Num. of authors dummies	Yes	Yes
State dummies	Yes	Yes
Subfield dummies	Yes	Yes

Table A6: Given name initials and US/China born Chinese

First name initials	(1) Distribution: Chinese names in US	(2) Distribution: Chinese names in China	(3) Distribution: Anglo-Saxon names in US	(4) Ratio (2)/(3)	(5) Chinese in US: US born	(6) Chinese in US: China Born
A	0.020	0.010	0.053	0.19	0.565	0.435
B	0.026	0.029	0.040	0.72	0.159	0.841
C	0.072	0.051	0.060	0.85	0.410	0.590
D	0.037	0.036	0.091	0.39	0.234	0.766
E	0.011	0.004	0.035	0.12	0.649	0.351
F	0.023	0.025	0.014	1.78	0.158	0.842
G	0.030	0.039	0.042	0.94	0.032	0.968
H	0.071	0.072	0.023	3.21	0.199	0.801
I	0.004	0.001	0.007	0.11	0.835	0.165
J	0.110	0.105	0.145	0.73	0.243	0.757
K	0.031	0.022	0.042	0.52	0.404	0.596
L	0.058	0.061	0.040	1.52	0.187	0.813
M	0.043	0.033	0.091	0.36	0.364	0.636
N	0.011	0.008	0.016	0.50	0.378	0.622
O	0.001	0.001	0.002	0.24	0.639	0.361
P	0.027	0.018	0.043	0.42	0.437	0.563
Q	0.024	0.035	0.000	126.09	0.000	1.000
R	0.025	0.020	0.098	0.21	0.333	0.667
S	0.068	0.060	0.068	0.88	0.295	0.705
T	0.033	0.020	0.042	0.46	0.492	0.508
U	0.000	0.000	0.001	0.16	0.610	0.390
V	0.004	0.001	0.007	0.13	0.767	0.233
W	0.052	0.053	0.038	1.39	0.208	0.792
X	0.065	0.096	0.000	804.65	0.000	1.000
Y	0.107	0.123	0.002	68.01	0.000	1.000
Z	0.046	0.079	0.001	80.58	0.000	1.000