Online Appendix for:
Do Return Requirements Increase International Knowledge Diffusion?

Appendix A: Data Appendix

Below is a description of the steps followed when constructing the dataset used in the paper.

Identifying Fulbrights

The names of Fulbrights were obtained from volumes of *Foreign Fulbright Fellows: Directory of Students* published annually by the Institute for International Education (IIE) from 1993 to 1996.

Data from Dissertations

The *ProQuest Dissertations and Theses* database is a database of almost all dissertations filed at over 700 U.S. universities. We obtained information from this database on students' full names, advisors, fields of study, Ph.D. completion dates, and undergraduate institution and/or country of birth. Starting in the1990's, ProQuest began publishing online the full text of the first 24 pages of the dissertation.

Matching Procedure

We first entered data from the IIE volumes on the Fulbright Student's name, graduate institution, field of study, and country of origin. Then, we searched for these students in the *ProQuest* database (described below) to find their date of graduation (for those who completed their studies) and advisor name. For those Fulbrights successfully completing their programs, we then performed searches on Google, Google Scholar, LinkedIn, and/or Web of Science to obtain as much information as possible on all of the student's post-Ph.D. locations and affiliations. Research assistants were instructed to limit search time to 20 minutes. If a student was not found at all on the web within 20 minutes, the searcher moved on to the next name.

For the students found on the web, we then searched for controls. We searched for controls obtaining Ph.D.s in the same year, with the same advisor, at the same institution as the Fulbright. If this step failed (i.e. if there were no foreign students with the same advisor graduating in same year), we looked for a student with the same advisor graduating within 3 years of the Fulbright. When choosing controls, we alternated students graduating before the Fulbright with those graduating after the Fulbright so that on average controls graduate at the same time as Fulbrights. If this step failed, we choose a control graduating in the same year in the same field of study (e.g. Biochemistry) at the same university. Finally if we had found more than one control with the same Ph.D. year, institution, and advisor (with the same year, institution and field), we chose the one from the same general region (although not from the same country).

The PhD Institutions of the Fulbright/control pairs in our data set are listed in Table A-1.

Identifying Country of Origin

For schools listing prior degrees or biographical information in the dissertation, we used this information to infer the student's country of origin, as described in the next paragraphs. For schools that did not list prior degrees, if we found a potential control student, we looked at their dissertation acknowledgements to see if they identified the country of origin. Failing this, we looked them up on the web using the methods described below under "Finding Location". Several universities require students to list biographical information in the front matter of the dissertation.[1] Table A-2 lists the universities that include biographical information and that conferred PhDs on Fulbright/control pairs in our sample. At some universities, the information includes a full biographical sketch (e.g., Ohio State, NC State), but in most cases, the information is limited to a list of previous degrees. Figures A-1 and A-2 present examples of this information drawn from dissertations filed at the University of Illinois and the Ohio State University.

The biographical information contained in these dissertations can be used to identify the country of origin of the student. Under the assumption that most students attend undergraduate programs in their country of origin, we treat the country of undergraduate degree as the country of origin. Using this information as a proxy for the nationality of the student will of course introduce some error, since not all students receiving undergraduate degrees do so in their country of origin. However, evidence from the NSF's *Survey of Earned Doctorates* suggests that the country of undergraduate degree is a very good proxy for the country of origin. For students completing doctorates in 2003 and 2004, when the *SED* listed the country of undergraduate degree, 84.9% of students obtained their undergraduate degree in the same country as their citizenship. This percentage was likely to be even higher a decade earlier. We also compared the counts of the number of doctoral recipients by country of origin, university and year computed from a *ProQuest* sample have a correlation of 0.948 with analogous counts obtained from the *SED*.

[1] The data on country of origin is only available beginning in the late 1990's when universities began submitting digital copies of dissertations to be posted on the web by ProQuest. However, by 1996 or 1997 almost all dissertations are available in digital format.

Publication Data

We obtained publication histories from *ISI's Web of Science*. Authors were identified using information on post-Ph.D. locations (collection procedure described below), authors' middle names, and fields of research. For each publication by an author, we obtained all information available on the publication record itself, including publication year, title, co-author names, author locations, counts of forward citations, publication source, abstract, specific field (for example, Marine & Freshwater Biology), and keywords.

It should be noted that our information on the number of forward citations received by an article includes self-citations.

The *ISI Web of Science* database does not cover every scientific journal published worldwide. At the time at which our publication data was collected (in 2008) it listed articles from 6,410 scientific journals.[2] Among Thomson's criteria for including a journal in the index are "The journal's basic publishing standards, its editorial content, the international diversity of its authorship, and the citation data associated with it." Journals must typically publish on-time, implying a substantial backlog of articles forthcoming. They must publish bibliographic information in English, and must include full bibliographic information for cited references and must list address information for each author. Thomson also looks for international diversity among contributing authors, but regionally focused journals are evaluated on the basis of their specific contribution to knowledge. The number of citations received by the journal is a key factor in evaluation for inclusion in the index, with preference going to highly cited journals or journals whose contributing authors are cited highly elsewhere.

The ISI selection procedure is designed to select the most relevant scientific journals, independent of the location of their editorial offices. Since such a large share of cutting-edge science research takes place in the U.S., there will inevitably be a high share of journals in this index based in the U.S. Journals that do not publish bibliographic information in English are less likely to be included, so articles written abroad and published in low-profile regional journals with limited readership beyond the region (as evidenced by a failure to publish bibliographic information in English) will be excluded from our data. As a result, our publication data should be viewed as information on scientists' participation in the international scientific community, rather than raw article counts. Still, the large number of journals included, and the special consideration given to regionally-focused journals means that most of the relevant journals in which our scientists publish

[2] See Thomson's *Journal Citation Report* 2008.

will be included. We examined the publication records of some of our scientists located outside the U.S., and found that even what might seem like relatively obscure journals (e.g. *Revista Chilena de Historia Natura, Revista Brasileira de Ciência do Solo*, *Acta Pharmacalogica Sinica*, etc.) were all included in the ISI index. While it is possible that ISI data is less comprehensive for articles published in non-Roman alphabets, it should be noted that only a very small number of scientists in our sample are located in Asian countries (5.7% of person-year observations have a location in Asia). Furthermore, these are scientists who began their careers in the United States and are thus likely to continue publishing in English-language journals.

To verify more rigorously that our sample of publications is not biased towards finding articles by U.S.-based researchers, we performed the following test. We had a research assistant collect data on the number of articles listed on scientists' C.V.s and the number of articles we obtained from ISI. We computed the share of a scientist's articles from the C.V. that were listed in the ISI database, and performed a t-test of difference in means between scientists outside the U.S. and those inside the U.S. The average share of articles found on Web of Science was 0.705 for those in the U.S. and 0.651 for those outside the U.S. We cannot reject the hypothesis of no difference in means (with a t-statistic of 0.788 and p-value of 0.433 for a two-tailed test). We thus do not feel that a systematic U.S. bias is introduced by restricting our attention to journals included in the ISI index.

We made sure to collect information on Fulbright and Control publications at the same time, ideally on the same day. We did this to avoid biasing the data to include more pubs and cites for one of the groups because they were collected later and had more time to appear in the database.

Identifying Current Location

To identify the locations of sample scientists at a point in time, research assistants searched on the web using Google, Google Scholar, LinkedIn and Web of Science (when the publications identified the location of our scientist). We interpolated location when the person was observed in the same institution and country before *and* after the interpolated date.

To ensure the comparability of our Fulbright and control samples, we also compared the types of online information used to identify locations for the two samples. These sources are very diverse and are similar for controls and Fulbrights, as shown in Appendix Figure A-3. This chart shows the distribution of sources for our location data. The most common is an online CV, with 28.7% of observations on controls' locations, and 29.7% of Fulbrights', coming from this source. Other common sources are LinkedIn, with 22.3% of control observations and 17.7% of Fulbright observations, faculty websites or web bios with 14.7% and 17.7% of observations, respectively, and publications, with 14.6% and 13.9% of observations respectively.

Moreover, average publications are not different for those academics who post and do not post their CVs or LinkedIn profiles and this is true both for those in the US and abroad, as shown in Table A-3. This Table displays regressions of our publication variables on a) the CV dummy alone, or b) the CV dummy, foreign location dummy, and the interaction of these two variables. These regressions show that there is no statistically significant difference in publication counts between those who do and do not post their CVs online. They also show that this is true both for those in the US and for those abroad.

Match Statistics

Table A-4 shows the distribution of Fulbrights and controls by PhD year. For exactly 125, or exactly 50.2%, of pairs, the Fulbright and control graduated in the same year, and 77.1% graduated within one year of each other. Only 2% of the pairs graduated more than 3 years apart, with the maximum time difference at 7 years for one pair (the control graduated in 1992 and the Fulbright in 1999).

Table A-5 shows the distribution of Fulbrights and controls by country. For 29.7% of the pairs, the control and Fulbright come from the same region of origin. Europeans paired with other Europeans make up 17.2% of the sample. Among those with discordant regions of origin, the most common pattern was Latin American Fulbrights matched with Asian controls (representing 17.7% of the sample). The second most common pairing was Latin American Fulbrights with European controls (15.3% of the sample), followed by European Fulbrights paired with Asian controls (7.6% of the sample). As we note in the paper, many non-Fulbrights from Mexico or other Latin American companies displayed evidence in the thesis acknowledgements of having been subsidized by their governments, and therefore not used as controls.

There are 79 pairs, or 31.7% of the sample, in which the advisor is the same for both members. The broadly-defined field is the same for 82.7% of the pairs. In the large majority of cases, the scientists in "different" fields did research in the same broad area, but were classified in different interdisciplinary fields, e.g. one student in "environment" and the other in "earth/air/ocean", or one in biological sciences and the other in agricultural sciences.

Match Comparisons to NSF's Survey of Earned Doctorates

To determine whether our data collection methodology resulted in a dataset representative of the population of foreign-born U.S. Ph.D. recipients, we compared our data to the Survey of Earned

Doctorates (SED), a database compiled by the National Science Foundation. This latter database contains detailed information on demographic and educational characteristics and post-degree plans for doctoral recipients at American universities at the time of receipt of degree between 1957 and 2008. All individuals receiving doctorates from accredited American universities are requested to complete the survey, and the response rate to the most recent wave of the survey was 93%.[3] The SED asks respondents about their countries of origin, doctoral institution, plans for post-graduate study or employment, and migration plans.

In 1996 and 1997, the SED also collected information about sources of funding support for graduate studies and specifically asked doctoral recipients whether they had Fulbright funding (reported in the SED as the SRCE, or sources of support, series of variables, code 61). In 1998 and later, this question was no longer asked. We therefore use only 1996 and 1997 data from both our XXX database and the SED to compare Fulbrights and non-Fulbright foreign students.

To match our XXX data, we restricted the SED to doctoral recipients in the natural sciences on temporary visas. We exclude non-Fulbright survey respondents who received some other form of funding support from a foreign government, because these individuals were excluded from the XXX sample because they typically receive J-1 visas like Fulbrights. However, results were comparable when these individuals were included. Finally, we also restricted the SED sample to those who had data on intended sector of work, gender and plans to migrate.

The sample of Fulbrights for which SED information on post-degree employment plans for location and job-sector is available is small for 1996-97 PhD recipients (88 in total). When we restrict this to the Fulbrights graduating from institutions covered by the XXX sample (See Table A-1), this number falls to 66. This compares to 59 Fulbright PhD recipients in the years 1996-97 in the XXX data whose location and employer type we have for the PhD year. We would not expect these two numbers to be exactly the same[4] but are gratified that the numbers are the same order of magnitude. There are 67 control PhD recipients in the years 1996-97 with non-missing observations on location and job type in the year after PhD in the XXX data.

We first compared the sector of first job. For the SED, we created a categorical variable capturing whether the person intended to work in academia, government or non-profit organization, or private organization post-PhD. This variable is based on information in the PDEMPLOY and PDOCPLAN fields in the SED. The latter captures whether the respondent planned to be employed

---

[3] http://www.nsf.gov/statistics/srvydoctorates/, accessed 5/23/13

[4] After all, in the SED, not everyone has employment plans at PhD-receipt and the SED has a less than 100% response rate; in the XXX data, we do not have location and employment information for all Fulbrights for the year of PhD itself, etc.

or further their education (e.g. in a post-doctoral position) and the former categorizes the type of planned employment. For the XXX data, we identified the job sector of the first job in which we observe each person.

The comparisons for Fulbrights and non-Fulbright foreign students are shown in Table A-6. Despite the small sample size, the SED results are remarkably similar to what we observe in the XXX sample for the first year post-PhD. Thus our sampling and data construction methodology appears to map the job sector of the universe of doctoral recipients remarkably closely.

We also created a dummy variable equal to 1 if the SED respondent reported that they planned to leave the United States after receiving the doctorate and compared this to the first post-PhD location observed for the XXX sample. (Table A-6) Here, there is a difference between the SED and the XXX sample but it is similar for controls and Fulbrights. Thus, the percentage of respondents to the SED who report intentions to stay in the US exceeds the percentage of scientists observed in the US in the following year. This suggests that the intentions expressed in the SED may be overly optimistic in terms of staying in the US.[5]

Finally, we compared the percent female in Table A-6 and found similar percentages despite the small samples.

In sum, we conclude that our data collection method does not appear to have biased our sample towards including individuals in particular sectors of employment or with a particular gender, when compared to a census of doctoral recipients. The distribution of characteristics of individuals in our sample is remarkably similar to the distribution of characteristics in the population.

Publication Comparisons to NSF's SDR

To investigate whether our sampling methodology and matching procedure led us to identify PhDs who were either more or less productive in research than random scientists, we compared our publication data to the Survey of Doctorate Recipients (SDR), a database compiled by the National Science Foundation. The SDR follows a random sample of U.S. doctorate recipients *who remain in the US* as their careers unfold, with survey waves typically every 2 years. We therefore compared publications from the SDR respondents to publications of the XXX sample for those who remain in the US.

The SDR publication data available is the *self-reported* number of published articles, and is asked only in selected waves, covering a different number of years in each of these waves. Since the

[5] In work comparing the SED to the longitudinal NSF SDR for a project on postdocs, Kahn and Ginther (2014) found that some SED respondents were also overly optimistic about the likelihood of finding a post-doc.

XXX data uses only *Web of Science* publications rather than the less selective "published articles in refereed professional journals", the SDR is likely to report more publications than the Web of Science. We have thus collected publication counts from the CVs of the scientists in our sample who have remained in the US continuously since graduation (and for whom a CV was available). By using counts drawn from the CV, we ensure that we capture all the articles that would be reported on a survey like the SDR, rather than the selected articles covered by the Web of Science.

In Table A-7, we report results from the 2001 SDR (which covered up to 6 years of publications) and the 2003 SDR (which covered 2 years of publications) for the PhD years and sub-sample we chose.[6] The SDR sub-sample we chose includes only temporary residents in STEM fields to allow comparability with the XXX data. Also, it is limited to cohorts with the particular PhD years (shown in Table) that allow matching with the XXX data.

Table A-7 indicates that the article counts for the scientists in our sample are similar to the article counts reported by the SDR, for each survey wave. In no survey wave is the difference between our sample's mean publications and the mean publications in the SDR statistically significant.

In sum, we conclude that there is no evidence that our data collection methodology has substantially biased the number of publications of those observed living in the US relative to a random sample of similar scientists.

[6] Publication data are available in the 1995 SDR wave as well which covered PhDs primarily before 1994 and therefore would have very few XXX observations. Similarly, publications are available in the 2008 SDR wave and covered 5 years of data but, since most of the XXX sample received PhDs 1994-2000. Using this SDR wave would to prohibitively small XXX comparison samples, with less than 10 people.

Table A-1: Institutions Granting Doctorates to Scientists in our Sample

Arizona State University	University of California at Berkeley
Auburn University	University of California at Davis
Boston University	University of California at Irvine
Brandeis University	University of California at Los Angeles
Brown University	University of California at Riverside
California Institute of Technology	University of California at San Diego
Case Western University	University of California at Santa Barbara
City University of New York	University of Chicago
Clemson University	University of Cincinnati
Columbia University	University of Colorado at Boulder
Cornell University	University of Colorado Health Sciences Center
Georgia Institute of Technology	University of Connecticut
Harvard University	University of Florida
Indiana University Bloomington	University of Georgia
Iowa State University	University of Illinois at Chicago Circle
Johns Hopkins University	University of Illinois at Urbana-Champagne
Louisiana State University	University of Iowa
Massachusetts Institute of Technology	University of Kansas
Michigan State University	University of Maine
Mount Sinai School of Medicine	University of Maryland at College Park
New York University	University of Massachusetts at Amherst
North Carolina State	University of Michigan
Northeastern University (MA)	University of Minnesota
Notre Dame University	University of Missouri-Saint Louis
Ohio State University	University of Nebraska (Lincoln)
Oklahoma State University	University of Nevada
Oregon State University	University of Oregon
Penn State University	University of Pittsburgh
Princeton University	University of Rhode Island, RI
Purdue University	University of Rochester
Rensselaer Polytechnic Institute	University of Rochester, NY
Rice University	University of South Carolina
Rutgers University	University of Southern California
Stanford	University of Texas at Austin
Stevens Institute	University of Texas at Dallas
SUNY Albany	University of Washington
SUNY Buffalo	University of Wisconsin at Madison
SUNY Stony Brook	University of Wyoming
Temple University, PA	Virginia Polytech
Texas A & M University	Washington State University
Texas Tech University	Wayne State University
University of Arizona	Worcester Polytechnic Institute, MA
University of Arkansas	Yale University

Table A-2: Universities Listing Biographic Information in the Thesis in our Sample

BOSTON U
CORNELL U
LOUISIANA STATE U
NC STATE
OH STATE
OK STATE
TEXAS A&M
U ARKANSAS
U CALIFORNIA
U CINCINATTI
U COLORADO
U CONNECTICUT
U FLORIDA
U ILLINOIS
U MAINE
U MASSACHUSETTS
U MISSOURI
U NEVADA
U OREGON
U PITTSBURGH
U SOUTH CAROLINA

Table A-3: Publication Variables regressed on a Dummy 1 for Online CV or LinkedIn Profiles

	Total Pubs	Total (fwd) Citations
CV	0.0380	0.0347
	(0.105)	(0.147)
CV	-0.0302	0.0108
	(0.133)	(0.167)
Foreign location	-0.480***	-0.690***
	(0.157)	(0.230)
Foreign location *CV	0.0280	-0.109
	(0.205)	(0.306)

Note: Poisson Regression of article variables on controls for year, year of PhD, log of home country GDP 5 years prior to graduation, gender, and scientific field. "CV" is a dummy equal to 1 if information on the scientist's location was derived from an online CV or LinkedIn profile. "Foreign location" is a dummy equal to 1 if the scientist was located outside the US in the previous year. The insignificant coefficients on "CV" in the first panel show that scientists who post online CV's do not have more publications than other scientists in the sample, for any measure of publication. The second panel shows that this is true for scientists in the US as well as those abroad.

Table A-4: Distribution of Controls and Fulbrights, by year of Ph.D.

Year of PhD	Controls	Fulbrights	Total
1991	1	0	1
1992	2	0	2
1993	9	5	14
1994	13	19	32
1995	13	23	36
1996	33	29	62
1997	44	35	79
1998	40	41	81
1999	32	34	66
2000	29	23	52
2001	13	22	35
2002	9	10	19
2003	7	6	13
2004	2	1	3
2005	2	1	3
Total	249	249	489

Table A-5: Distribution of Controls and Fulbrights by Country of Origin

Country of Origin	Controls	Fulbrights	Total	Country of Origin	Controls	Fulbrights	Total
Argentina	3	4	7	Kenya	0	2	2
Armenia	1	0	1	Korea	8	0	8
Australia	0	4	4	Lesotho	0	1	1
Austria	3	3	6	Lithuania	0	1	1
Bangladesh	2	0	2	Macedonia	1	0	1
Belgium	1	3	4	Malawi	1	1	2
Bolivia	0	1	1	Malaysia	1	0	1
Botswana	0	1	1	Mexico	9	96	105
Brazil	11	0	11	Morocco	0	2	2
Bulgaria	1	0	1	Netherlands	4	5	9
Canada	9	0	9	Nigeria	1	0	1
Chile	3	0	3	Norway	2	6	8
China	18	0	18	Pakistan	2	0	2
Colombia	5	8	13	Panama	1	1	2
Costa Rica	0	3	3	Peru	2	2	4
Cote D'Ivoire	1	3	4	Philippines	3	2	5
Croatia	1	1	2	Poland	1	1	2
Cyprus	1	0	1	Portugal	2	19	21
Czech Republic	3	1	4	Romania	10	0	10
Denmark	2	4	6	Russia	1	0	1
Ecuador	1	0	1	Singapore	0	1	1
Egypt	2	0	2	Solomon Islands	0	7	7
Ethiopia	2	2	4	South Africa	6	7	13
Finland	2	5	7	Spain	1	0	1
France	2	0	2	Sri Lanka	1	0	1
Germany	10	0	10	Swaziland	2	3	5
Ghana	0	2	2	Sweden	3	1	4
Greece	4	7	11	Switzerland	7	0	7
Guatemala	1	2	3	Taiwan	1	1	2
Haiti	0	1	1	Tanzania	5	5	10
Hungary	8	2	10	Thailand	0	2	2
Iceland	2	7	9	Togo	1	1	2
India	25	0	25	Trinidad & Tobago	13	1	14
Indonesia	4	0	4	Turkey	2	4	6
Iran	1	0	1	UK	1	2	3
Iraq	1	0	1	Uganda	5	0	5
Ireland	2	1	3	Ukraine	2	1	3
Israel	3	6	9	Venezuela	3	0	3
Italy	5	3	8	Yugoslavia	1	0	1
Japan	5	0	5	Zimbabwe	0	2	2
Jordan	1	0	1	Total	249	249	498

Table A-6: Post-Grad Employment Plans in the SED and our Sample

Full Sample		SED Restricted Sample		XXX Sample	
Controls (XXX) or SED Non-Fulbright Sample					
% planning work in:		% planning work in:		% working in:	
acad	61.0%	acad	60.0%	acad	62.1%
govt	16.8%	govt	16.3%	govt	16.7%
priv	22.2%	priv	23.7%	Priv	21.2%
% female	19.9%	% female	19.3%	% female	24.7%
% planning to go abroad	24.9%	% planning to go abroad	25.4%	% located abroad	32.8%
Fulbrights					
% planning work in:		% planning work in:		% working in:	
acad	63.1%	acad	61.5%	acad	69.5%
govt	15.5%	govt	15.4%	govt	16.9%
priv	21.3%	priv	23.1%	Priv	13.6%
% female	23.9%	% female	27.3%	% female	21.9%
% planning to go abroad	36.4%	% planning to go abroad	36.4%	% located abroad	69.5%

Table A-6: Average self-reported publications since Ph.D., in the SDR and our XXX sample

PhD Year	XXX MEAN	SDR MEAN	Difference	t-stat for test of difference
1996	4.25	5.11	-0.9	-0.908
1997	5.5	4.87	0.6	0.399
1998	4.87	5	-0.1	-0.118
1999	3.71	5.18	-1.5	-1.230
2000	4.63	5.51	-0.9	-0.567
2001	6.29	5.45	0.8	0.440

Figure A-1

ALGORITHMS AND ARCHITECTURES FOR SOFT-DECODING REED-SOLOMON CODES

BY

ARSHAD AHMED

B.E., Regional Engineering College, Trichy, 1998
M.E., Indian Institute of Science, Bangalore, 2000

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Electrical Engineering
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2006

Urbana, Illinois

Figure A-2

VITA

January 31, 1973	Born – Da-An, Jilin Province, China
September 1989 - July 1993	Bachelor of Science in Electrical Engineering, Nanjing University of Science and Technology, Nanjing, China
September 1993 - April 1996	Master of Science in Electrical Engineering, Nanjing University of Science and Technology, Nanjing, China
Septmeber 2002 – present	Ph.D student, Analog VLSI Laboratory, Department of Electrical and Computer Engineering, the Ohio State University, Columbus, Ohio
Since June 2006	RFIC design engineer, Freescale Semiconductor Inc., Boca Raton, Florida

PUBLICATIONS

Research Publications

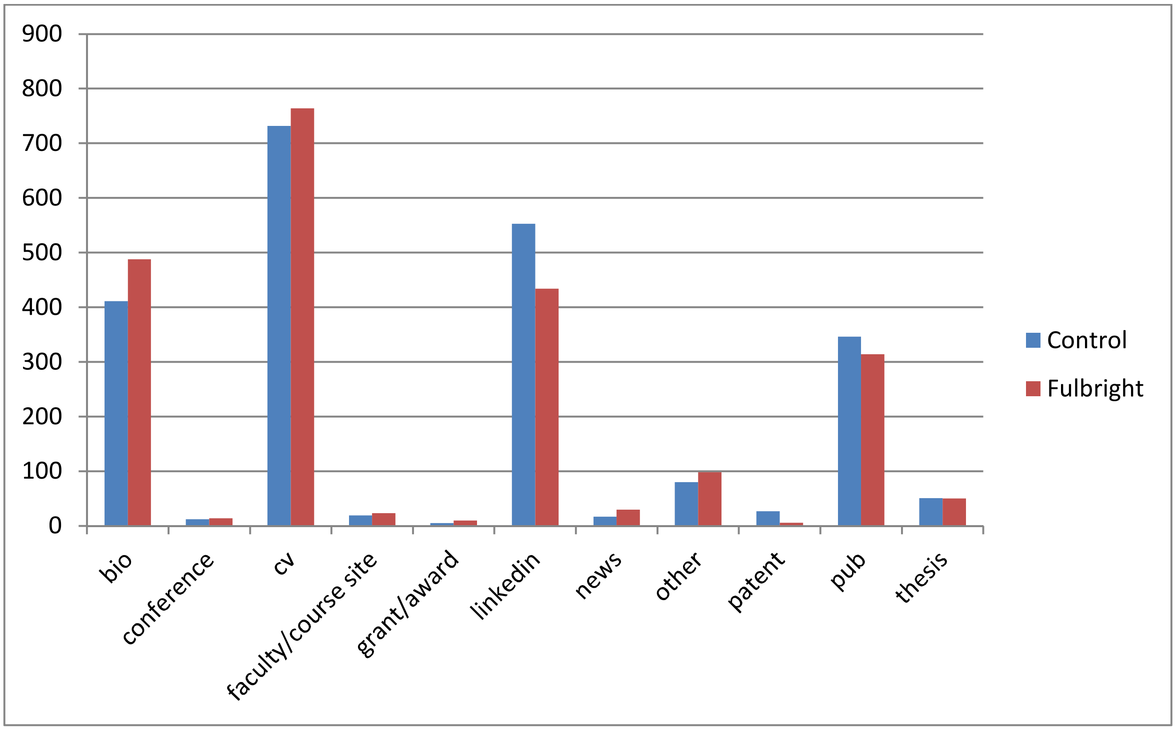
P. Zhang, and M. Ismail "A New RF Front-End and Frequency Synthesizer Architecture for 3.1–10.6 GHz MB-OFDM UWB Receivers", *Proc. 48th Midwest Symposium on Circuit and System*, vol.2, pp.1119–1122, August 2005.

C. Garuda, X. Cui, P. Lin, S. Doo, P. Zhang, and M. Ismail "A 3–5 GHz Fully Differential CMOS LNA with Dual-gain Mode for Wireless UWB Applications", *Proc. 48th Midwest Symposium on Circuit and System*, vol.1, pp.790–793, August 2005.

Y. Yu, L. Bu, S. Shen, B. Jalali-Farahani, G. Ghiaasi, P. Zhang, and M. Ismail "A 1.8V Fully Integrated Dual-band VCO for Zero-IF WiMAX/WLNA Receiver in 0.18μm CMOS", *Proc. 48th Midwest Symposium on Circuit and System*, vol.1, pp.187–190, August 2005.

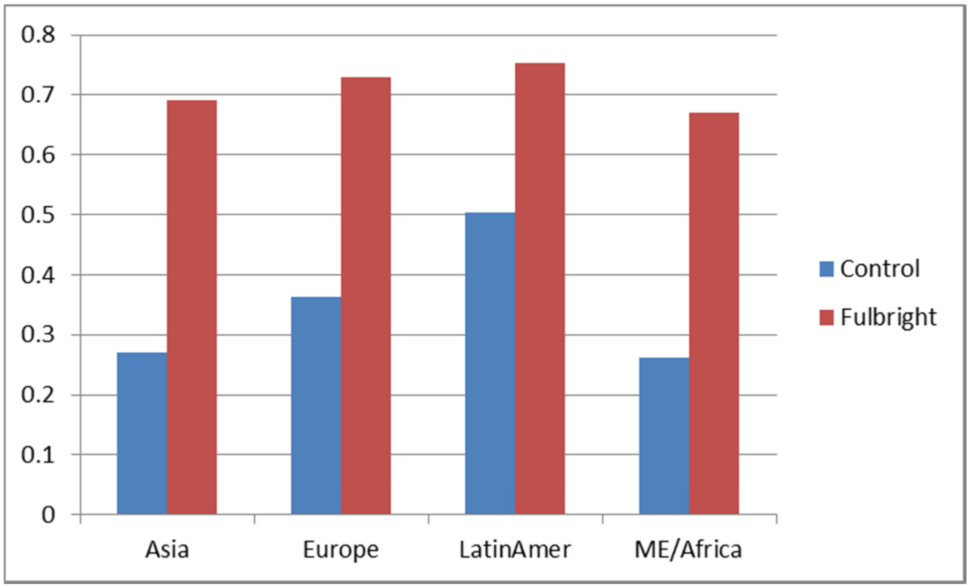
vii

Figure A-3: Sources of Location Information, by Source Type



The sources of location data are categorized as follows:
bio: short biography
conference: conference website or proceedings listing institutional affiliation
cv: curriculum vitae
faculty/course site: website for a course taught by the scientist at a particular institution or faculty website at an institution
grant/award: documentation of a grant or award listing affiliation
linkedin: LinkedIn page listing employment history with dates and locations
news: news articles or press releases mentioning the scientist's affiliation or location
patent: patent document listing location
pub: journal article listing authors' affiliations
thesis: student thesis listing the scientist as an advisor

Figure A-4: % of post-Ph.D. years spent abroad, by region of origin



Note: “Europe” includes Australia, Canada, and Israel.

Online Appendix B: Additional Robustness Checks

B1. Details on Alternative Matching Methods

Our first alternative matching approach involves matching each Fulbright to the non-Fulbright who is the closest to them in terms of the predicted probability of being a Fulbright.[7] We estimate a logit model in which the dependent variable is the Fulbright dummy and explanatory variables are the control variables previously used (Ph.D. program rank, gender, year of Ph.D., field, GDP per capita of home country, pre-graduation publications) plus dummies for region of origin. The pseudo-R^2 of this regression was 0.206. Based on this logit model, we calculate the fitted values of this regression as the predicted probability of being a Fulbright based on observable characteristics. Each of our Fulbrights is then matched to the control student in the same broad scientific field with the closest predicted probability of being a Fulbright.[8] We think of this as a way of collapsing the observable characteristics correlated with Fulbright status into an index that can then be used to identify the most similar control to the Fulbright in question. There are 201 Fulbrights and 91 controls in the final dataset, with the typical control appearing 2.3 times in the re-matched dataset. The first columns in Table B1 show the average values of the covariates among Fulbrights and controls matched on the predicted probability of being a Fulbright. All covariates are balanced (there is no statistically significant difference in the means of these covariates after matching).

The second alternative matching procedure approach is inspired by the Coarsened Exact Matching (CEM) procedure used by Azoulay et al. (2010) following Iacus, King, and Porro (2008). In this non-parametric procedure, the econometrician selects a set of covariates for which balance between treated and control groups is desired. These covariates are then recoded to group together into strata observations with "substantively similar" values of the variables in question (Iacus et al., 2011). For each treated observation (where treatment in this case is Fulbright status), a matching control observation is drawn from its stratum. If there is no matching control in a stratum, the observation is dropped.

We create strata based on the scientist's field of study, the GDP per capita of the home country (5 years before Ph.D.), the number of first-authored articles written while in grad school, the year of graduation, and the rank of the Ph.D. program. It would be impossible to match exactly on all of these characteristics while still obtaining a dataset with enough observations to ensure statistical

---

[7] This is similar to propensity score matching, but the matching is with respect to the propensity to be a Fulbright (the instrument) rather than with respect to the propensity to be abroad (the treatment).
[8] First, we "trim" the dataset in the manner of Crump et al. (2009) by eliminating the bottom 10% and the top 10% of the distribution of predicted probabilities.

power. Therefore, we "coarsen" the distribution of all covariates but the field of study, on which we require an exact match.[9] We then randomly select a control to match to each Fulbright from the same stratum in which the Fulbright is found. Controls may be sampled more than once. Given our relatively small universe of 498 controls and Fulbrights, we must use fairly coarse bins to avoid dropping large numbers of individuals due to failed matches. Specifically, we divide the covariates into strata based on whether the value of a covariate is above or below the median of that variable in the dataset, retaining 402 individuals while dropping 96 (21 Fulbrights and 75 controls) for whom a match was not found.[10] The average control appears 1.422 times in the new dataset. Table B1 (right columns) shows the average values of the covariates among Fulbrights and controls matched using CEM. Again, all covariates are balanced except the log of the GDP per capita of the home country five years before graduation, which is included as a control variable in many of our specifications (see text).

The results described in Table B2 and Table B3 are broadly consistent with the main results of the paper. The propensity score matching results are somewhat more variable, because the number of controls in the sample is dramatically reduced, increasing the potential noise in the data.

Table B1
Comparisons of Fulbright and Control Samples – Alternative Matching Methods

	CEM			**matching on P(Fulbright)**		
	Mean for Controls (N = 174)	Mean for Fulbrights (N = 228)	p-value of t-test of difference in means	Mean for Controls (N = 91)	Mean for Fulbrights (N = 209)	p-value of t-test of difference in means
LAGFORLOC	0.358	0.719	0	0.306	0.72	0
Real GDP per cap of hc 5 yrs prior to PhD	8.868	8.931	0.016	8.972	8.951	0.854
Pre-PhD 1st authored pubs	1.206	1.184	0.897	1.44	1.349	0.707
Log(university rank)	0.665	0.748	0.466	0.49	0.638	0.29
Year of PhD	1997.839	1997.811	0.911	1997.675	1997.842	0.572

[9]Note that field is already considerably coarsened, as we have grouped together narrower fields into broad subject areas.
[10]If, instead, we coarsen each covariate into terciles, we end up with only 179 individuals in the dataset.

Table B2: Robustness of forward citation results to alternative matching approaches

	(1)	(2)	(3)	(4)	(5)	(6)	(1)	(2)	(3)	(4)	(5)	(6)
	CEM, Forward cites from home (Table 2)						**Propensity score, Forward cites from home (Table 2)**					
Fulbright	-0.006	0.605**					-0.0245	0.335				
	(0.268)	(0.251)					(0.330)	(0.318)				
Fulbright from country >median articles per capita in field			0.467	0.563*					0.221	0.22		
			(0.401)	(0.311)					(0.480)	(0.411)		
Fulbright from country <median articles per capita in field			-0.387	0.713*					-0.105	0.753**		
			(0.398)	(0.378)					(0.474)	(0.327)		
Fulbright from country >median cites per article in field					0.494						0.15	
					-0.329						(0.375)	
Fulbright from country <median cites per article in field					0.710**						0.678**	
					(0.302)						(0.344)	
Fulbright from country >75th pctile GDPpc						0.536*						0.605
						(0.309)						(0.474)
Fulbright from country <75th pctile GDPpc						0.769**						0.237
						(0.383)						(0.376)
	CEM, Forward cites from USA (Table 3)						**Propensity score, Forward cites from USA (Table 3)**					
Fulbright	0.0267	-0.032					-0.259	-0.159				
	(0.243)	(0.139)					(0.293)	(0.179)				
Fulbright from country >median articles per capita in field			0.776**	0.16					0.288	0.131		
			(0.327)	(0.142)					(0.408)	(0.178)		
Fulbright from country <median articles per capita in field			-0.900***	-0.328					-1.049***	-0.695**		
			(0.315)	(0.208)					(0.336)	(0.289)		
Fulbright from country >median cites per article in field					0.167						0.196	
					(0.158)						(0.176)	
Fulbright from country <median cites per article in field					-0.224						-0.545**	
					(0.205)						(0.251)	
Fulbright from country >75th pctile GDPpc						0.132						0.0868
						(0.150)						(0.183)
Fulbright from country <75th pctile GDPpc						-0.28						-0.616**
						(0.220)						(0.297)

Table B3: Robustness of backward citation results to alternative matching approaches

	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
	CEM, Backward cites to home (Table 6)				**Propensity score, Backward cites to home (Table 6)**			
Fulbright	0.738***				0.593**			
	(0.260)				(0.279)			
Fulbright from country >median articles per capita in field		0.646**				0.686*		
		(0.302)				(0.353)		
Fulbright from country <median articles per capita in field		0.603				0.336		
		(0.431)				(0.369)		
Fulbright from country >median cites per article in field			0.682**				0.308	
			(0.329)				(0.310)	
Fulbright from country <median cites per article in field			0.515				1.194***	
			(0.346)				(0.345)	
Fulbright from country >75th pctile GDPpc				0.551*				0.975**
				(0.328)				(0.403)
Fulbright from country <75th pctile GDPpc				0.820*				-0.0106
				(0.423)				(0.347)
	CEM, Backward cites to US (Table 7)				**Propensity score, Backward cites to US (Table 7)**			
Fulbright	0.0189				0.0809			
	(0.063)				(0.083)			
Fulbright from country >median articles per capita in field		0.079				0.121		
		(0.084)				(0.115)		
Fulbright from country <median articles per capita in field		-0.0482				0.0367		
		(0.089)				(0.123)		
Fulbright from country >median cites per article in field			0.0845				0.275**	
			(0.091)				(0.111)	
Fulbright from country <median cites per article in field			-0.0448				-0.102	
			(0.080)				(0.089)	
Fulbright from country >75th pctile GDPpc				0.0307				0.0242
				(0.096)				(0.117)
Fulbright from country <75th pctile GDPpc				0.014				0.179
				(0.092)				(0.135)

B2. Robustness of results on US forward citations described in Table 3

Table 3 shows that there is no statistically significant difference (at the 5% level) between forward citations from US researchers to articles by Fulbrights and articles by controls. Table B4 presents robustness checks with alternative specifications for citations originating in the US. The coefficient on Fulbrights from low-science countries is not qualitatively different from the result in Column 4 of Table 3 when we add sector dummies, consider only global journals, or exclude those in agricultural/environmental fields (Columns 1, 2, 3 respectively). However, the coefficient falls in magnitude and becomes clearly insignificant in the final three specifications, those that in various ways control for the article's or scientist's quality. We conclude that the marginally significant negative effect for Fulbrights from low-science countries in Column 4 of Table 3 may just reflect the fact that these scientists have fewer very highly-cited articles.

Citations from the US to Fulbrights from *high*-science countries remain insignificant in all these robustness checks, except for Columns 5 and 6, where the coefficient becomes significantly *negative*. These results suggest that, while on average there is no significant reduction in the number of citations coming from the US for articles by scientists from high-science countries who are subject to return requirements, this finding may largely be driven by relatively few highly-cited Fulbrights from high-science home countries. Without these (or controlling for them), Fulbrights from these high-science countries get fewer US citations than either comparable controls from high-science home countries or Fulbrights from low-science countries.

Table B4: Citations from US: Robustness

	(1)	(2)	(3)	(4)	(5)	(6)
	Adding sector dummies	Only Global journals	Only Global, Excluding Agri/Enviro	Controlling for journal impact	Controlling for other citations	<90th pctile of cites
Fulbright from country >median articles per capita in field	0.00879	-0.0773	-0.0760	-0.00915	-0.398***	-0.364***
	(0.144)	(0.144)	(0.159)	(0.150)	(0.115)	(0.124)
Fulbright from country <median articles per capita in field	-0.287*	-0.305*	-0.343	-0.217	-0.140	0.00872
	(0.167)	(0.174)	(0.218)	(0.160)	(0.0980)	(0.114)
Home country <median articles per capita in field	-0.0911	-0.288*	-0.273	-0.301*	-0.235**	-0.299***
	(0.197)	(0.169)	(0.203)	(0.171)	(0.0975)	(0.108)
ln US publications in field in citing year	0.613***	0.564***	0.676***	0.426**	0.126	0.0172
	(0.227)	(0.213)	(0.261)	(0.216)	(0.108)	(0.123)
ln Publications by scientist in cited year	1.456***	1.441***	1.472***	1.482***	0.238***	0.843***
	(0.0845)	(0.100)	(0.118)	(0.0935)	(0.0760)	(0.101)
Cites to home-country articles in pregrad pubs	-0.00315**	-0.00474***	-0.00461***	-0.00505***	-0.00394***	0.00486
	(0.00158)	(0.00146)	(0.00171)	(0.00150)	(0.000636)	(0.0110)
Share of pubs in high-impact journals				0.889***		
				(0.142)		
ln Citations from other countries					1.029***	
					(0.0364)	
Sector dummies	Y	N	N	N	N	N
Observations	39,816	30,270	39,816	39,816	39,816	35,754

Poisson regression coefficients with robust standard errors, clustered by scientist, in parentheses. *** p<0.01, ** p<0.05, * p<0.1
All specifications include controls for PhD rank as well as dummies for gender, field, citing year, year of PhD and citation lag (T-t).. Coefficients of these controls available upon request.

References

Azoulay, Pierre, Joshua Graff-Zivin and Jialan Wang (2010), "Superstar Extinction", *The Quarterly Journal of Economics* 125 (May): 515-548

Crump, Richard, V. Joseph Hotz, Guido Imbens and Oscar Mitnik (2009), "Nonparametric Tests for Treatment Effect Heterogeneity", *Review of Economics and Statistics* Vol 90(3): 389-405.

Iacus, Stefano M., Gary King, Giuseppe Porro G. (2012). "Causal Inference Without Balance Checking: Coarsened Exact Matching." *Political Analysis* ,20 (Winter), 1-24.

Kahn, Shulamit and Donna Ginther (2014) *The Postdoc Rat Race*. Working paper, Boston University.

Thompson Reuters. *Journal Citation Reports*. http://thomsonreuters.com/journal-citation-reports/