How Important is U.S. Location for Research in Science?

Shulamit Kahn*

Megan J. MacGarvie**

Abstract

This paper asks whether scientists located outside the U.S. are at a disadvantage when it comes to research productivity. We use a dataset of foreign-born U.S.-educated scientists that allows us to exploit exogenous variation in post-Ph.D. location induced by visa status. We thus are able to compare students who were required by law to leave the U.S. after the completion of their doctoral studies with similar students who were allowed to remain in the U.S. We assess whether a student who left the U.S. has more or fewer publications and/or citations when compared to an otherwise similar control student. Instrumenting for location using visa status and accounting for the current country's real GDP per capita, we find that the negative relationship between non-U.S. location and research output is present and large for countries with low income per capita but completely eliminated when the researcher is located in a country in the highest deciles of GDP per capita. This suggests that a scientist exogenously located in a country at the top of the income distribution can expect to be as productive in research as he or she would be in the U.S.

---

* Boston University ** Boston University and NBER. email: skahn@bu.edu, mmacgarv@bu.edu. We thank the Institute for International Education and Jerry Murphy for the Fulbright directories. We thank Olesya Baker, Gina Brandeis, Maria Burtseva, Eleanor Cooper, Gerardo Gomez-Ruano, T.J. Hanes, Dana Schulsinger, and Vanessa Wong for excellent research assistance. We thank Richard Freeman, Rodrigo Canales, Jenny Hunt, Carlos Serrano, and Paula Stephan for helpful comments. This project was funded by National Science Foundation Grant SBE-0738371.

The United States produces more doctorates in science and engineering (S&E) than any other country, is home to a disproportionate share of top scientists (NSF *Science and Engineering Indicators* 2010, Zucker and Darby 2007) and has many of the most highly rated universities in the world.[1] However, the share of S&E doctoral degrees produced outside the U.S. has grown in recent years (NSF *Science and Engineering Indicators* 2010), and some countries have increased efforts to attract star scientists.[2] In light of the increasing "globalization" of science, many observers have questioned whether U.S. dominance in science and engineering can be expected to continue (Wadhwa et al. 2007). The answer is likely to hinge on whether the U.S. research environment offers inherent advantages to scientists whose productivity would fall if they located elsewhere, or instead whether the U.S. has merely succeeded in attracting a large number of exceptional scientists who would be productive in any location. Geographic proximity to a high concentration of top scientists that creates large knowledge spillovers, top journals and scientific conferences headquartered in the U.S., extensive financial resources devoted to R&D, and a U.S. culture of academic competitiveness may all contribute to a special advantage in science for the U.S.. On the other hand, these advantages may no longer be determinative because of advances in communication technology and increased investment of other countries in science and as a result, other countries may be able to woo scientists to relocate and the U.S. could lose its "edge" in science and engineering research and potentially other types of innovation (Furman, Porter and Stern 2002).

This paper directly addresses the question of whether U.S.-based scientists are more productive than those located elsewhere. Specifically, we compare U.S.-educated, foreign-born scientists who stayed in the U.S. to those who went abroad post-Ph.D., in terms of their research output (measured by the number and prestige of publications, first and last authorship), diffusion and impacts on science (measured by the number of citations to their articles – *forward citations*), and connection to cutting-edge science (measured by the median lag of articles cited in their publications – *backward citations*).

A first look at the research records of a sample of 488 foreign-born scientists who received U.S. Ph.D.'s between 1991 and 2005 (Table 1) clearly indicates that compared to those located outside the U.S., the U.S.-located scientists produce more publications and more publications in "high-impact" journals (defined below) each year, with the difference most stark between the U.S.

---

[1]A Chinese ranking of the world's top Universities places the U.S. as having 15 and 17 of the top 20 universities in the world in natural sciences/math and engineering/computer science respectively. (Shanghai Jiao Tong University 2008).The ranking is based on Nobel laureates and Fields medals prize winners, citations and publications. We thank Bound, Turner and Walsh (2006) for identifying this source.

[2]For instance, the Canada Research Chairs program and the Australian Research Council's Federation Fellowships offer incentives to attract researchers to these countries.

and lower income countries but still large for higher income ones. Later tables also show that these U.S.-located scientists' articles receive significantly more citations and cite more recent literature.

However, comparisons of scientists inside the U.S. with those outside are plagued by unobserved heterogeneity among scientists and endogeneity of location choices. For instance, better scientists may be more likely to receive U.S. job offers, to desire to remain within the U.S. research community and/or to face a wide wage differential between locating in the U.S. v. abroad. Therefore, a naïve comparison of the publications of scientists in different locations would not isolate the *impact* of location on productivity as distinct from unobserved individual productivity and preferences.

In the present paper, we address this problem by constructing a new dataset carefully crafted to isolate the impact of U.S. location by exploiting exogenous variation in post-Ph.D. location induced by visa status. The research productivity of 244 foreign-born Ph.D. recipients who came to the U.S. through the Foreign Fulbright Fellowship program and legally must leave the U.S. upon the completion of their studies is compared to that of a control sample of 244 foreign-born Ph.D. recipients who received their Ph.D.s in the same university, field, time period (and, if possible, with the same advisor) but who had no visa requirements to leave the U.S. Our data is unique in being the only data set of which we are aware that tracks the career progression of individual U.S.-trained Ph.D. scientists, whether they leave the U.S. or not.[3]

Although we constructed our Fulbright and control samples to be very similar in ability and background, because there may remain differences in inherent research potential between the two groups, we also control for additional background factors including country of origin and pre-Ph.D. publication record. In addition, we investigate whether our results change under different hypotheses about bias or using alternative matching methods.

To summarize our results, we find that on average being outside the U.S. leads to fewer total, last-authored and high-impact publications and to fewer citations to high-impact publications. Further, we find that the negative impact of being outside the U.S. is present and large for those in countries with low levels of GDP per capita but seems to be eliminated for rich countries. Also, scientists in lower-income countries cite a less recently published literature.

These findings have implications about the global dispersion of research during the coming decades, and particularly whether the U.S. is likely to retain its current research edge if present trends continue. In the paper's conclusion, we discuss possible implications and kinds of government policies that might alter the future path of geographical dispersion of research. Before proceeding to a

[3]One can obtain information on foreign-born scientists who remain in the U.S. from the NSF's SESTAT database. Also, Michael G. Finn's (2007) research provides valuable information on the stay rates of Ph.D.s. of foreign origin.

description of our data, our empirical results and discussion of the implications of these results, we first address in the following section *why* U.S. location may directly affect research productivity.

II. Why U.S. Location May be Important

Why might the U.S. be a more productive location for scientific research? First, geographic proximity to other scientists is thought to enhance knowledge diffusion and collaboration. In those geographic areas within the U.S. with greater stocks of knowledge (as measured by past articles, patent applications of scientists working there, the presence of star scientists or highly-ranked academic institutions), there are more new publications, patents and innovations by both private companies and academics. However, the research on proximity's advantages has struggled to establish this as the *impact* of location working through geographic knowledge spillovers from nearby scientists, rather than mere correlation. The existence of spillovers is suggested by the increased likelihood of researchers, within the same geographical area, to collaborate across institutions and sectors and to cite each others' articles and patents, by the tendency of new firms to locate near universities active in that field, and by the impact of exogenous changes in universities' R&D funding on geographically-close companies.[4] The causal impact of top scientists in prestigious departments on new assistant professors has been isolated by instrumenting prestige with demand and supply factors affecting the academic market at the time of the initial placements (Oyer 2006 and Stephan and Levin 1992).

Given evidence that geographic closeness to excellent universities and scientists spills over to others' research productivity, the high concentration of top scientists in the U.S. gives all scientists located within the U.S. an advantage to the extent that these knowledge spillovers are more likely to occur within a country than between countries. Moreover, for U.S.-educated scientists, geographic closeness is likely to foster increased collaboration with Ph.D. advisors and fellow students. Consequently, it seems likely that foreign-born recipients of U.S. doctorates who return to home countries will be less productive than those who remain in the U.S. Of course, the same reasoning applies to scientists living far from universities and research centers but *within the U.S.* Nevertheless, distances within the U.S. are on average small compared to the thousands of miles of water and land that separate the U.S. from all but its few nearest neighbors.

Geographic propinquity *per se* is not the only avenue through which scientists can benefit from U.S. location. (1) Well-funded American universities and research institutes can devote

[4]See Agrawal, Cockburn and McHale (2006), Audretsch and Feldman (1996), Audretsch, Lehmann and Warning (2005), Jaffe (1989), Jaffe, Trajtenberg and Henderson (1993), Zucker, Darby and Brewer (1998), Zucker and Darby (2006) and Zucker, Darby, Furner, Liu and Ma (2007). However, Orlando (2004) and Thompson and Fox-Kean (2005) have contested the strength of some of this evidence.

considerable financial resources to increasingly expensive research laboratories and equipment. (2) Many top journals are located in the U.S. and have mainly American editors and reviewers. (3) Scientists located in the U.S. face lower costs of participating in the many well-attended conferences, seminars and meetings in the U.S. where they can network and present their research. (4) Culture itself may be a factor. Many U.S. universities place a high value on successful research – indicated by prestigious publications and grant awards – and consequently U.S. academia is often characterized by a high degree of competitiveness. (5) Common institutional structures within U.S. academia and research centers makes collaboration within the U.S. easier. (6) Finally, there may be a wider range of jobs in basic scientific research available in the U.S.

Moreover, initial career advantage tends to lead to later advantage in academia.[5] Consequently, students who leave the U.S. post-Ph.D. for visa reasons even for just a few years are likely to have their research careers permanently affected.

Several countervailing forces might have mitigated many of the advantages enjoyed by U.S. researchers in the last half of the twentieth century. Advances in communications technology and reductions in the cost of international travel have reduced geographic barriers to knowledge diffusion and to long-distance collaboration in science. Over the past decade particularly, other governments have made the development of stronger research capabilities a national priority at a time when U.S. policies may have deterred some scientific explorations. These countries are expanding and improving their doctoral-level educational capabilities, partially by successfully attracting more star scientists. The U.S. share of S&E Ph.D.'s being awarded is decreasing, with Freeman (2006) documenting that in the past two decades, the major Asian Ph.D.-producing countries went from graduating fewer than half the number of U.S. S&E Ph.D.'s to graduating more, and somewhat less dramatically, EU countries went from graduating fewer to more S&E Ph.D.'s than the U.S. Between 2003 and 2007, countries other than the U.S. slightly expanded their share of the top 100 universities, although they have not made new inroads into the top 20 (Shanghai Jiao Tong University 2003, 2007). Finally, private sector S&E jobs abroad have also become more widely available as multinationals increase their non-U.S. employment of research scientists and U.S. companies off-shore some high-level S&E jobs to foreign-owned companies.[6]

These factors might be the cause of the increasing collaboration and citations observed recently across state and international borders and decreasing impacts of being in a top university on

[5]See Stephan (1996), Oyer(2006).
[6]See Freeman (2006) and Kim, Morse and Zingales (2006).

research productivity.[7] They may also account for the increasing propensity of U.S.-trained highly skilled immigrants (including top scientists) to return to their home countries – dubbed by Annalee Saxenian as a "brain circulation" replacing "brain drain"[8] – and for increasing effects of R&D spending across international borders.[9] Below, we investigate whether these trends have indeed erased any productivity advantage for scientists working within the U.S.

III. Empirical Approach and Data Set

A. The Foreign Fulbright Program as an Instrument

As noted earlier, comparisons of U.S. and foreign scientists' research output will inevitably be plagued by selection bias since scientists' locations are likely to be influenced by unobserved characteristics correlated with productivity. Specifically, the most productive foreign-born U.S.-educated scientists may be most likely to stay in the U.S. The strategy we use to identify the separate impact of location on productivity is to identify pairs of foreign-born U.S. S&E (excluding social science) Ph.D. recipients from the same department in the same university graduating during the same period (and, whenever possible, with the same advisor and from the same general region) – one of whom is a Fulbright Fellow with a J-1 student visa required by law to leave the United States for at least two years after finishing his/her doctorate, and one of whom faced no such restrictions.

For Fulbright status to be a useful instrument, we must first establish that far more Fulbright scholars actually do leave the U.S. than do other foreigners studying in the U.S. The requirement to leave the country after the completion of studies is quite stringent. It is possible to apply for a waiver of the foreign residency requirement if a student falls into one of several very restrictive and quite rare categories.[10] Also, Fulbright recipients may delay their departure for two years of a post-doc and/or for up to three years of "occupational or practical training" (OPT) on-the-job immediately following the completion of their studies. Thus, in principle, a Foreign Fulbright recipient could remain in the U.S. for up to 5 years following the receipt of Ph.D. before having to leave the country.

[7] See Adams et al.(2005), Agrawal, Kapur and McHale(2007), Kerr(2008), Kim, Morse and Zingales(2006), Singh(2005).
[8] See Saxenian (2002), Zucker and Darby (2006).
[9] Kerr (2002).
[10] One route is for students to ask their country of origin to file a "no-objection" statement, although this approach is almost never considered grounds for waiving the foreign residency for Fulbrights. (Conversation with BU ISSO January 2008). Waivers may also be obtained if an "Interested Government Agency" files a request on behalf of the student, stating that departure of the student will be detrimental to its interest and that of the public. Our conversations with experts suggest that these waivers are obtained only in rare circumstances. Medical doctors may obtain a waiver if they agree to practice in a region of the U.S. with a shortage of health care professionals. Waivers can be given if "an exchange visitor believes that he or she will be persecuted based on his/her race, religion, or political opinion if he/she were to return to his/her home country." Finally, applications for waivers may be filed on the basis of "Exceptional hardship to a United States citizen (or legal permanent resident) spouse or child of an exchange visitor." The State department warns, "Please note that mere separation from family is not considered to be sufficient to establish exceptional hardship." http://travel.state.gov/visa/temp/info/info_1288.html (accessed Feb. 17, 2008). Economists can remain in the U.S. if they become part of international organizations like the World Bank, an avenue not available for most natural scientists.

Moreover, the Fulbright-subsidized Ph.D. can apply for a work visa and return to the U.S. after two years spent in their home country. These two years need not even be 730 consecutive days, but could be a combination of summers and/or semester-long visits abroad while spending the rest of the time in a U.S. post-doc or in OPT.

The enforcement of these rules is sufficiently stringent that almost all foreign Fulbright Ph.D. recipients leave the U.S. for some period post-Ph.D. Only 12.3 percent of our Fulbright sample appear to have remained in the U.S. continuously and 23.4 percent appear never to have been in their home country post-Ph.D. and thus to not have fulfilled their home country residency requirement, although even they could have fulfilled the requirement in short segments that we did not observe. For the other 76.6 percent of the Fulbright students in our sample, we were able to find evidence that they did spend some time in their home country after receiving their Ph.D.s, compared to only 36.1 percent of our control group of US-educated foreign-origin non-Fulbrights.

We observe our sample of 244 Fulbright scholars for a total of 2,042 person-years post- Ph.D. 75.8% of these years are spent outside the U.S and 63.2% in the home country itself. In contrast, the 244 controls spent only 34.2% of their 2,108 observed person-years outside the U.S. and 27.7% in their home countries. This U.S. stay rate of approximately 66% for control students is nearly identical to the average stay rate estimated in a much larger sample by Finn (2007), who found that 67% of foreign students who received their doctorates in 1998 (close to the average Ph.D. year in our sample) were observed in the U.S. in 2003. A large difference between Fulbrights and controls in their tendency to be outside the U.S. exists for each region of origin. In sum, the Fulbright instrument is strongly correlated with the endogenous variable foreign location.

In order for Fulbright status to be a legitimate instrument, we must also establish that our Fulbright sample is similar to our control group with respect to *potential* research productivity and proclivity at graduation. Our matching of each Fulbright with a control was done precisely in order to create two groups with identical research potential. In order to convince readers that we have done so, we discuss the construction of our control sample in some depth. Additional details on sample construction are available in an online Data Appendix.[11]

For each Fulbright-funded Ph.D., we used the *ProQuest Dissertations and Theses* database to obtain information on the year of graduation and advisor and to identify a “control” student of foreign origin who did *not* have post-Ph.D. location restrictions and who graduated from the *same* program in

[11] Appendix materials available at http://people.bu.edu/skahn/

the *same* year and, whenever such a student existed, with the same advisor.[12] Since students who receive substantial funding from their home country's government often are required to return for some period, we searched for evidence of foreign governmental funding and excluded the student as a control if we found any.

When several potential control students were identified for a single Fulbright fellow, we chose the student who came from the same region, or more generally, from countries similar (particularly in GDP per capita) to the countries represented in the Fulbright sample. However, the distribution of students across countries in the treatment and control groups, while similar, is not identical. There are several reasons for this. First, the distribution of Fulbrights across countries is affected by past and present political factors. Second, because many students from certain countries receive government funding, we were less likely to select controls from those countries.[13]

We then searched for the Fulbrights' and controls' locations since their Ph.D. receipt on Google, Google Scholar, LinkedIn, and/or Web of Science, the combination of which allowed us to find both academics and non-academics.[14] If people were found in the same location several years apart, we assumed that they had been at that location continuously. If there were still years missing location, we extrapolated it for one year. If, after this process, we could not identify the location of either the Fulbright or the control for at least half of the years since Ph.D., we dropped that pair. Since it tends to be easier to find academics' locations on the web than others', we no doubt under-sampled non-academics. Similarly, since it tends to be easier to find successful people on the web – be they academics or non-academics – we no doubt under-sampled the less successful. However, we believe this under-sampling applies equally to Fulbrights and controls.

In sum, the match between treated (i.e. Fulbright) and control students was made with the goal of choosing controls that are as similar as possible along the characteristics relevant to our study.

---

[12]In cases where there was no control student with the same advisor in the same year, we identified a student with the same advisor graduating within 3 years before or after the Fulbright. If no students met the latter criteria, we chose a student graduating in the same year in the same major field, but with a different advisor.

[13]There are considerably more Fulbrights from Mexico, Portugal, and South American countries and more controls from China, Germany, Italy and Turkey. (Complete listing available in online Appendix.) There are no Fulbrights from China or India in our sample so we tried to avoid sampling controls from these countries, but when a suitable control could not be found from another country, we allowed students of Chinese and Indian origin in the sample. The reason that there are many Fulbrights from Mexico but no controls is that most of those without Fulbright fellowships are subsidized by their governments.

Note that for countries with a large number of Fulbrights, it is preferable *not* to have controls from the same country since it becomes more likely that those who did not receive Fulbright's were of lower quality.

[14]Many academics had C.V.'s posted on the web which included their location and their previous locations. Non-academics were more likely to be found on Linked In, conference or meeting programs, alumni associations, local news articles or civic/religious organizations websites. One person was even located via a DUI arrest. We made sure that the person we located had more than just their name in common with the student we knew (e.g. the Ph.D. location or a previous employer might be mentioned.) Note that although we had many people's CVs, for consistency we did not base our publication counts on them.

The criteria we used for matching were based on our priors about the characteristics that are most relevant for future research output (institution, advisor/field, date of graduation and, where possible, region of origin). Statistics on the closeness of the matches are included in the online Data Appendix.

It is possible that, due to the inherent difficulties of finding controls that are exactly identical to the Fulbrights along every dimension except visa status, there may be differences between controls and Fulbrights that introduce bias. The direction of the bias is not obvious. Countries generally have their own committees that award Fulbright Fellowships based on their own selection criteria. While we assume that academic excellence is one of these criteria – which would lead to Fulbrights having more research potential – the committees might avoid funding the most promising students if they are believed to be less likely to spend their careers in their home country – which would lead to Fulbrights with less research potential. Moreover, many excellent students may not accept Fulbright Fellowships if they have strong preferences to remain in the U.S. post-Ph.D. and/or can afford to avoid funding that restricts their futures. Similarly, it is possible that we may over-sample or under-sample those living abroad who are less successful at publishing. For instance, people living outside the U.S. who are less successful in publishing may nevertheless publicize their accomplishments (e.g. by posting their CV) in order to be noticed abroad; or conversely, people living outside the U.S. who are less successful at publishing because they are no longer associated with academic circles may be less likely to publicize their accomplishments in ways accessible to our web-searching.

To the extent that U.S. universities can observe the differences between students, the university admissions procedure should ensure that the Fulbrights and non-Fulbrights they admit to any specific department have equivalent abilities. However, future research preferences may not be entirely observable to U.S. university admissions and/or U.S. departments may lower their standards for well-funded graduate students. Finally, Fulbright Fellows may not apply to the university best matched to their abilities. In many countries, Fulbright commissions guide Fellows towards particular U.S. universities, sometimes influenced by the availability of supplementary fellowship funding from the university and/or the lower tuition costs of public universities.[15]

We have done several things in order to remove and/or evaluate possible biases due to differing inherent research potential of Fulbrights and controls. First, we include explanatory variables likely to be correlated with or to capture research ability and proclivity. One such variable is the Ph.D. institution's rank. Since this is identical within the pair, inclusion of this variable only serves to increase the explanatory power *across* pairs. Second, in most specifications we control for the GDP per capita of the home country at the beginning of graduate school.

---

[15]Conversation with IIE representative, June 2009.

Third, in some specifications, we also include as control variables measures of students' research output while in graduate school which we believe to be a good proxy for inherent ability. Including these pre-grad publication variables may over-control in the sense that at least some of the Fulbright-control differences in pregrad publications may also be a result of being a Fulbright. For instance, if Fulbrights believe that they will return home to a non-research job, they may be less committed to getting their Ph.D. research published. On the other hand, if Fulbrights are more concerned about having opportunities to leave their home country after two (or more) years of residence, they may feel they need stronger credentials. Poisson regressions of students' pre-graduation research output variables on the Fulbright dummy, region of origin, field and Ph.D. year indicate a 22.5% lower rate of total publications among Fulbrights level not significant at conventional levels (p-value=.121) and an even less significant 12.7% lower rate among Fulbrights for first-authored publications (p-value=.314). We believe the first-authored pre-grad publication count to be a more direct measure of student's future research potential.[16]

We also do a variety of robustness checks. In Section IV.B., we estimate how biases of reasonable magnitudes would affect our estimates. In Sections IV.C and D, we report results of alternative specifications and alternative matching methods. Our main conclusions remain robust in these analyses.

B. Estimation Model

The basic estimating equations in this paper are:

$$(1)\ Y_{it} = \alpha + \beta\, LAGFORLOC_{it} + \gamma\, X_{it} + \varepsilon_{it}$$

$$(2)\ LAGFORLOC_{it} = \varphi + \xi\, \text{FULBRIGHT}_{i} + \psi X_{it} + \eta_{it}$$

where Y is one of nine measures of research output/diffusion, *LAGFORLOC* is whether or not the individual was in a non-U.S. location in the previous year, and X is a vector of exogenous control variables. Each observation is a person-year starting from the year after Ph.D. receipt and continuing through 2007. Half of the people in our sample are Fulbright Fellows in S&E Ph.D. programs listed in the 1993-1996 annual volumes of *Foreign Fulbright Fellows: Directory of Students*.[17] We included all Fulbright Fellows in these volumes who (1) received Ph.D.'s in S&E (excluding social science); (2) we located at least two years during the post-Ph.D. period; and (3) were in departments with at least one "matchable" control as described in the previous section.

The key right hand side variable is a dummy variable for whether the researcher was located

[16] Results are similar for Negative Binomial models for first authored publications, while the result has similar magnitude but greater significance for total publications. All pre-grad publication results are available from authors on request.

[17] 1993 was the first year these lists were published and included all Foreign Fulbrights *enrolled* in U.S. graduate programs in that year. Later cohorts would not have had sufficient post-Ph.D. careers.

in the U.S. the previous year. We have lagged this variable one year because of the typical time in science between when research is performed and when it is published.[18]

The data set includes only those years for which the lagged foreign location of the scientist is known, starting from the year after their Ph.D. graduation year through 2007. This leaves us with 4,150 observations.

C. Measuring Research Output

The dependent variables we model were taken from information on the Fulbright and Control Ph.D.s' publication histories from *ISI's Web of Science*.[19] From the *Web of Science*, we obtained information for the following publication-related variables:

Publication counts: The number of articles each year on which the scientist is a contributing author. This may be a noisy measure of research output when articles have many authors.

First-authored publication counts: The number of articles each year on which the scientist is the first author. In science, the first author is the major contributor to the research.

Last-authored publication counts: The number of articles each year on which the scientist is the last author. In science, typically the last author will be the person running the lab, who is often the Principal Investigator (PI) on the research grant funding the research. This variable is an indicator of the author's ability to secure research funding.

Publications in high-impact journals: The number of each year's publications in the top 50% journals *in that field* as ranked by ISI's impact factors (as of 2007). We made this measure field-specific because different fields have very different conventions about citations.[20]

Forward citation counts: The total number of citations received as of 2008 *by articles published each year*, which proxy a publication's impact on scholarship. We model citations to total, first-authored, and last-authored publications, and to publications in high-impact journals.[21]

Median citation lag: The median difference between the articles' backward citations and its publication date. The longer the lag, the less likely the article has been based on the most current science. Analyses of citation lags are limited to person-years with at least one publication.

---

[18]We also tried a two year lag with qualitatively similar results (available on request), but with larger standard errors because the number of observations were restricted. When more than one lag was included in a single equation, the separate coefficients were typically insignificant because FORLOC is highly serially correlated for each person.

[19]Authors were matched to publications using information on post-Ph.D. locations, authors' middle names, fields of research, co-authors on other work etc.

[20]List of high-impact journals available upon request.

[21]Due to the extreme skewness of their distributions, citation counts are winsorized at the 99th percentile. For instance, the maximum citation count was 1012 citations and we truncate this variable at 174.The few people with the highest counts of publications and citations were part of large collaborative physics labs where they appeared on all publications of the lab. Even for these people, the number of first and last-authored publications were measures of their specific contributions and were therefore not winsorized. Results obtained using raw publication and citation counts were qualitatively similar to the ones we report here.

Table 2 displays the publication and citation variables categorized both by present residence – U.S. or not – and by Fulbright status. The data by present location confirm our expectations. Ph.D. scientists in the U.S. do publish more articles, are more highly cited, and have longer (backward) citation lags. The differences are substantial. However, dividing the sample by Fulbright status tell a somewhat different story. Although the control scientists are much more likely to be living in the U.S., Table 2 shows that differences between controls' and Fulbrights' publications and citations average only about 41% of the U.S./non-U.S. spread.

D. Control Variables[22]

As explained above, although the Fulbrights and controls were matched to be observationally identical, we also include control variables to account for any remaining differences that may exist between Fulbrights and controls. The control variables include:

Ranking of Ph.D. institution: We include the (log of the) 1995 relative ranking of the U.S. Ph.D. institution (by field) from the National Research Council (Goldberger et al. 1995) to control for the quality of Ph.D. training. Lower rank signifies higher quality. Since rank is the same for Fulbright and control, this variable only increases the explanatory power across Fulbrights-control pairs.

Field dummies: Fields differ widely in the number of articles published a year and citing conventions. We categorized each student by the first field listed in their (*ProQuest*) dissertation record. We divided fields into the seven groups listed in Table 3.[23] Since the control and Fulbright were chosen from the same department, the distribution across fields of study should be exactly identical. There are small differences, however, since often the fields specified in *ProQuest* are quite narrowly defined and many dissertations list more than one field. Even students of the same advisor and department may list different first fields.

Dummies for calendar year and year of Ph.D receipt: Both variables are included in all specifications, with calendar year divided into 5 eras (93-96, 97-98, 99-00, 01-02 and 03-07) based on similar levels of average annual publications and citations.[24] There is a similar but not identical distribution across Ph.D. years between Fulbrights and controls, since the control was the closest

---

[22]Employment sector dummies – government, industry, academia – might pick up one reason that scientists in foreign locations are less productive, the scarcity of good academic jobs. However, sector is endogenous in that the best researchers move to countries with more academic jobs. In additional specifications (not shown), we included these dummies. This made no qualitative differences to our conclusions. We also experimented with including a dummy for students from countries with English as an official language which also did not change our results.

[23]Because of our small data set, we were unable to converge the instrumented model for most output variables for a more detailed set of field dummies. We experimented with different field groupings and qualitative results were not affected.

[24]Again, we were unable to converge the instrumented model for certain dependent variables (notably last-authored pubs and high-impact pubs) with a full set of year and Ph.D. year dummies. We *have* estimated the un-instrumented Poisson regressions with the full set of year and Ph.D. year dummies and did not find results to differ substantially from the results using the more restricted year and year from Ph.D. variables. This is likely due to the fact that our sample of controls and Fulbrights is approximately evenly balanced in terms of year and field characteristics.

available foreign student within three years of the Fulbright's Ph.D. (although the mean and median year of graduation are the same.) We also divide Ph.D. year into intervals as follows: pre-1995, 96-97, 98-99, 00-02, and post-2002.

Gender: We obtained data on the gender of the scientist using information from web searches (e.g. photographs, the use of personal pronouns in web bios), using a web-based algorithm for identifying the probable genders of given names when no other information was available.[25]

GDP per capita of home country 5 years before Ph.D. receipt: The log of the GDP per capita of the scientist's country of origin before they enter their Ph.D. program may affect the quality of pre-doctoral training or the average financial resources available for the student's doctoral education. On the other hand, this measure is highly correlated (ρ=.956) with the home country's present GDP per capita. As a result, when the analysis includes current GDP, we do not include this variable as a control, as explained in the results section.

Number of articles, first-authored articles, and first-authored high-impact articles published during graduate school: [26] In some specifications, the number of pre-graduation publications is included. These variables may measure individual-specific variation in past research productivity to proxy for inherent research potential, although as discussed above, they may be endogenous. We extend the "while-in-doctoral-program" period through the year after completion of the doctorate, because these articles are very likely to reflect dissertation research rather than new work performed post-graduation. Note that first-authored articles are more prevalent during the Ph.D. year than later. In fact, for the average student with any pre-grad publications, 60% of the articles published during this graduate school time were first-authored, probably publications from their thesis work for whom the Ph.D. student was the primary author.

IV. Empirical Results

A. Estimation and Results

Our principal results are found in Table 4. We estimate publications and citations using a count-data instrumental variables model similar to Poisson regression developed by Mullahy (1997) and estimated via GMM.[27] Standard errors are clustered by scientist. Angrist (2001) has shown that the Mullahy model gives a consistent estimate of the local average treatment effect (LATE) in a

---

[25]The gender-guessing program is found at: http://www.gpeters.com/names/baby-names.php.

[26]We also experimented with including pre-doctorate high-impact and last-authored publications, but they never had a significant effect on later research, perhaps because there were so few of them.

[27]Mullahy (1997) has shown that an I.V. Poisson model with unobserved heterogeneity that enters additively in the exponential mean function will be inconsistent when errors are assumed to be additive. Mullahy proposes a transformation of the model with multiplicative errors that is not subject to this problem. We used Stata's ivpois function to estimate these models, modified to allow for clustered standard errors. Similar results were obtained using a linear 2SLS model.

model with a binary instrument, endogenous treatment variable and no covariates. Median citation lags are estimated using 2SLS. Table 4 presents only the coefficients of lagged foreign location. Coefficients of other control variables are as expected. Full estimates are available on request from the authors.

Our main instrument for foreign location is Fulbright status. In Panel A of Table 4, we present results with a single location variable *LAGFORLOC,* a single instrument Fulbright status, and without pre-Ph.D. publications. The F-statistic measuring the power of Fulbright status in predicting *LAGFORLOC* is high – 125.09 – indicating a strong instrument. First stage regressions for the instrumented results are included as the Appendix Table.

The results in Panel A of Table 4 tells us that foreign location has a negative and significant impact on total publications. We are less certain about foreign location's impact on total citations, since its coefficient is not significant even at the 10% level, with a point estimate equivalent to a 51% difference.[28]

Limiting the analysis to first-authored publications i.e. those in which the scientist had *the* major role in the research lowers the coefficients on foreign location's impact on both publications and citations and renders them insignificant.

In many scientific fields, last-authorship signifies that the person ran the lab and was the PI who obtained the funding. We find that being outside the U.S. has a significant, large negative average effect on last-authored publications (-62%). However, citations to these articles are not significantly lower on average.

Finally, *LAGFORLOC* also has a large and significant impact on publications and citations *in high-impact journals*, at -76% for publications and -81%. There is no relationship between foreign location and the median backwards citation lag.

Controlling for pre-grad publications (Panel B) lowers point estimates and significance levels of foreign location's impact across the board. However, although the gap between U.S. and abroad narrows, it is still statistically significant for total publications (p-value=.08), last-authored publications (p-value=.04), high-impact publications (p-value=.02) and citations to high-impact publications (p-value=.052).

Overall, the results from Panels A and B suggest negative effects of being abroad on some measures of research output, fewer when conditioned on publications while in graduate school.

However, the impact of being outside the U.S. on scientists is likely to be heterogeneous. One

[28]Percentage effects are calculated from Poisson coefficients (β) as exp(β) -1.

factor that may affect the magnitude of any negative impact of being abroad is the wealth of the country in which the scientist is located, measured by its real GDP per capita. We expect that the research output of those who are in less developed countries with fewer resources for research would be most hurt by being outside the U.S. However, the GDP of the scientist's present foreign location will also be endogenous, with the most research-oriented Ph.D. recipients more likely to get jobs in richer countries with more resources devoted to science.

We differentiate the effect of being outside of the U.S. by GDP per capita in two different ways. In Panels C and D of Table 4, the two endogenous variables related to foreign location are *LAGFORLOC* alone and interacted with the GDP per capita of the *current* country (*LAGFORLOC X ln(lagged)GDP)*. The instruments that we use are Fulbright status and the lagged (log of) GDP per capita of the *home* country. The GDP per capita of the home country captures the differential effect of location for students *from* countries with different levels of income. Since it is based not on present country but on home country, it is clearly predetermined. Note that in this case, we have not controlled for the home country's GDP 5 years period to Ph.D. receipt since it is so highly correlated with present (lagged) GDP of one's country of origin. The F-statistics on these instruments are 74.06 (LAGFORLOC) and 79.38 (*LAGFORLOC X ln(lagged)GDP)*. The first-stage regression results are found in the Appendix Table columns 2 and 3. Table 4 Panels C and D present estimates of the impact of being abroad at four points in the distribution of GDP per capita: at the 25th, 50th, 75th and 90th percentiles, controlling and not controlling for pre-grad publications respectively.

The income per capita of the country in which the scientist is located does seem to mitigate the effects of being outside the U.S., with the point estimates of the impact of being abroad more negative in poorer countries than in richer ones for all research measures. Thus, once we instrument to account for selection bias induced by endogenous differences in foreign locations, there is no statistically significant negative difference between scientists in the richest countries (90th percentile of GDP per capita) and those in the U.S. on *any* of the measures of research output. In other words, the negative impact of foreign location is completely absent in countries with higher GDP per capita. In contrast, scientists in low GDP per capita countries *are* significantly negatively impacted by foreign location. When pre-grad publications are not controlled for, scientists from countries at the 25th percentile of the GDP per-capita distribution are significantly negatively impacted by foreign location in terms of *all* publication and citation measures (including backwards citations), while those at the 50th percentile of the GDP per-capita distribution are significantly negatively impacted for all output measures except citations to last-authored articles. After controlling for pre-grad publications, point estimates and significance of the impacts fall, remaining significant for the 25th percentile.

Panels E through H of Table 4 use a different approach to measure separate effects of being in rich v. poor countries. They divide the sample into two groups, the first comprising those scientists from richer countries (defined as in the top quartile of countries' real GDP per capita in the year of graduation) and the second comprising the rest of the world.[29] For those from rich countries, the effect of being abroad on every measure of scientific output is actually positive rather than negative, and even significantly so for two measures.[30] In contrast, for scientists originally from poor countries, the impact of being abroad is negative (except for cites to first-authored publications) and significantly so for many measures (with and without pregrad publication controls). Thus, the impact of being outside of the U.S. was significantly negative for all four specifications for last-authored publications, high-impact publications, cites to high impact publications and median citation lags and only consistently insignificant for first-authored publications and cites to those publications.

Finally, in results not included here[31], we estimated the main models without instrumenting. We expected un-instrumented coefficients to be more negative than the instrumented ones because we believed that those with a higher propensity to publish are more likely to stay in the U.S. We did find the un-instrumented results to be more significant in general than the instrumented ones (e.g. in Table 4 Panels A and B.) However, we found that the instrumented impact of being abroad is often larger (in percentage terms) than the un-instrumented one. In the separate rich/poor regressions comparable to Table 4 panels E-H, we found that the effect of being abroad in a rich country is much more negative in the un-instrumented ones, often significantly so. Thus, selection bias dominated OLS results for those from rich countries: those with a higher propensity to publish are more likely to remain in the US. However, for those from poorer countries, the instrumented results were uniformly more negative than the un-instrumented ones. One possible explanation is that the instrumented results for those from poorer countries are picking up the local average treatment effects (LATE) for the compliers, the Fulbrights who comply with the foreign residency requirements of their visas and therefore are in their home countries. In contrast, un-instrumented results combine Fulbrights abroad with controls abroad who have voluntarily chosen to return to their home countries (the always takers), probably because they have more opportunities there or fewer opportunities in the US.

To summarize the major results from this section, research output, the dissemination of this output and familiarity with recent literature suffers for scientists who leave the U.S. for low-GDP

[29]Note that these analyses are able to include the home country GDP per capita (5 years pre-Ph.D.) as a control variable because there is no other GDP variable in the estimation. First stage results in the Appendix Table columns 4 and 5.
[30]Cites to first-authored publications not controlling for pre-grad publications, and cites to last authored publications controlling for pregrad publications.
[31]Available upon request from authors.

countries even after selection bias is accounted for by instrumenting. The one exception may be first-authored publications and cites to these publications, although even for these there is some indication of a negative impact of being in the very poorest countries. In contrast, being outside the U.S. does not seem to impede research or diffusion for researchers located in countries with more wealth and resources, and indeed in a few cases being in these richer countries (outside the U.S.) may even boost some aspects of research. Results controlling for pregraduation publications are likely to be an underestimate of the impacts, but even these indicate quite negative impacts of being in poorer rich countries.

B. Sensitivity of Results to the Exclusion Restriction

Perhaps the most serious potential criticism of our instrument is the possibility that Fulbright status is correlated with a researcher's quality independent of location, meaning that the exclusion restriction is violated. While we have previously discussed the ways in which we have attempted to guard against this possibility through our sample construction and use of controls for pre-graduation output, we feel that the importance of this issue implies that its potential impact should be carefully considered. In this section, we provide some estimates of the potential effects of bias introduced by correlation between the dependent variable and the instrument. Our approach is similar to that of Angrist, Imbens and Rubin (1996) who examine the implications of a violation of the exclusion restriction in the relationship between civilian mortality and draft status. In the interest of simplicity, we assume a linear model.[32]

Let Y_i represent the research measure for person i (publications, citations), D_i be *LAGFORLOC* and Z_i be the Fulbright instrument (with both D_i and Z_i being dummies.) The coefficient in the simple OLS regression of Y_i on D_i is:

$$\beta_{ols} = Cov(Y_i,D_i)/Var(D_i)$$

The 2SLS equivalent of β_{ols} is:

$$\beta_{2sls} = Cov(Y_i,Z_i)/Cov(D_i,Z_i).$$

The exclusion restriction in our context implies that if a student was awarded and accepted a Fulbright Fellowship and was located in the U.S. post-Ph.D., that student would have had the same outcome (publications, cites) as if he had not been a Fulbright. Similarly, if the student was located abroad, he would have the same outcome as if he had not been A Fulbright but still located abroad.

If this exclusion restriction does not hold, the I.V. estimator will be biased. Suppose that the true model is:

[32] As mentioned earlier, results estimated using 2SLS are very similar to the results from the non-linear Mullahy model.

$$Y_i = \beta_0 + \beta_1 D_i + Q_i + \varepsilon_i$$

where Q_i is the unobserved quality of the student. Then, the 2SLS estimator will be:

$$\beta_{2sls} = Cov(D_i + Q_i, Z_i)/Cov(D_i, Z_i)$$

$$= \beta_1 + Cov(Q_i, Z_i)/Cov(D_i, Z_i)$$

This says that the IV estimate of the causal effect will be biased by the ratio of the covariance between quality and Fulbright status to the covariance of location and Fulbright status.[33] Assuming:

$$Q_i = \gamma_0 + \gamma_1 Z_i + \xi_i$$

$$D_i = \alpha_0 + \alpha_1 Z_i + \nu_i$$

then $$Cov(Q_i, Z_i) = \gamma_1 \, Var(Z_i)$$

$$Cov(D_i, Z_i) = \alpha_1 \, Var(Z_i)$$

Therefore:

$$\beta_{2sls} = \beta_1 + \gamma_1/\alpha_1$$

where γ_1 is the additional number of publications from the quality effect of being a Fulbright, and α_1 is the increase in the probability of being outside the U.S. from being a Fulbright. Note that even if γ_1 is non-zero – in other words the exclusion restriction is violated – the amount of bias will be small as long as the instrument is very strong (α_1 is large in absolute value).

Let us assume that the quality effect of being a Fulbright means the Fulbright is either over- or under-placed at his Ph.D. institution relative to controls. For example, consider the possibility that Fulbrights are over-placed because institutions are attracted to the funding that Fulbrights bring with them. This would imply that Fulbrights are admitted to programs that would not have admitted them if they were a control, and as a result, Fulbrights from a given institution can be expected to have lower productivity than controls from that institution, independent of where they are located after the completion of the Ph.D. program.

To determine the magnitude of the bias this would introduce, suppose that Fulbrights are typically admitted by programs that are either of higher or lower rank than the best program that would admit the Fulbright without the fellowship. By definition, controls are accurately matched with their Ph.D. institutions. Using regressions with quartile dummies, we find that moving from the first to the second quartile is associated with an average decline in research output of 10% (averaged across the 8 regressions for the different research output measures.)

We then perform sensitivity analysis in which we consider the possibility that Fulbrights on

[33] Angrist et al (1996) compute 2SLS estimates of the impact of veteran status on mortality, using draft lottery number as an instrument. The concern in their case is that the exclusion restriction is violated because draft numbers affect mortality through their influence on years of schooling. They use this method to estimate the amount of potential bias.

average have either higher or lower inherent research potential than controls. We consider two scenarios, and for each scenario, we compute $\beta_1 = \beta_{2sls} - \gamma_1/\alpha_1$. The two scenarios are:

1. Fulbrights are admitted to programs ranked 25% higher than the program that would admit them without the fellowship (negative bias γ_1 = -10% of the mean of the dependent variable. This would be equivalent to Fulbrights attending Stanford instead of the University of Maryland.)
2. Fulbrights are admitted to programs ranked 25% lower than the program that would admit them without the fellowship (positive bias γ_1 = +10% of the mean of the dependent variable).

In Table 5, we present the results of this sensitivity analysis, using the 2SLS estimates of the effect of foreign location at each of the 4 income levels discussed in the paper as well as the average impact.[34] We find that the qualitative results for scientists from countries of different income levels are the same. People in countries with the lowest income levels who live abroad are negatively impacted, significantly so both in the statistical and economic meaning of the word. People in the richest countries have small and mostly insignificant differences in research output, and point estimates are often positive. Of course, the size of the negative impact is larger if Fulbrights actually were better than their peers and is smaller than if Fulbrights actually were worse, although the differences in magnitudes due to bias are not large. The one qualitative difference that is worth noting is that if Fulbrights are actually 1 quartile "better", there is a 90.5% chance that even those in the highest income countries are less likely to publish high impact publications.

Our overall conclusion drawn from this sensitivity analysis is that, for reasonable amounts of potential bias, either positive or negative, the strength of the instrument dominates the bias and our results are qualitatively unchanged. We recognize that if Fulbrights and controls are more dissimilar than assumed here, we would have difficulty interpreting our results, especially for countries at the high end of the income distribution. However, given the steps we have taken to ensure comparability of students within the matched pairs and the control variables we have included to account for student heterogeneity, we feel that bias large enough to invalidate our results is unlikely.

C. Additional Robustness Tests

In this section we perform a variety of robustness checks. First, although we have clustered errors by individual to correct standard errors, we check whether we get similar results if we collapse the data to the scientist level (Panels A - D of Table 6). In this specification, the dependent variables are averaged over the years since receipt of Ph.D. and the key endogenous variable is the *share* of years spent abroad, instrumented by Fulbright status using 2SLS. The independent variables are the

[34]Not controlling for pre-grad publications.

same as in our main results described previously[35]. This specification ignores all time varying covariates, particularly how many years have passed since Ph.D., due to collapsing the data. Thus, we lose information about which years the person is observed (since each scientist was not necessarily observed each year since Ph.D.) Nevertheless, the results in Table 6 (Panels A-D) are very consistent with those obtained from the panel data, with large significant effects associated with spending more years abroad for scientists from poor countries and insignificant or positive effects of being abroad for scientists from rich countries. The key difference is that now the impact on last authored publications and on cites to those publications is only sometimes significant.[36]

It may be that the aspect of countries most important to the success of scientists located there is not their GDP but instead the commitment of the country to basic science and research. In panels E – H of Table 6, we divide countries based on the number of scientific articles per capita produced by the country (lagged, in logs) obtained from NSF (2008) as a proxy for this commitment. High article per capita countries are defined as countries whose articles-per-capita are above the 75th percentile. We find results very similar to the results in the previous panels A – D based on GDP per capita.

We also pursue two alternative matching approaches to correct for the possibility that Fulbrights and controls are fundamentally different. In our preferred matching procedure described in the main text of the paper, we match exactly on Ph.D. institution and field of study, and nearly exactly on year of graduation. However, it is possible that by choosing to match exactly on these covariates, we have neglected other factors that could introduce bias – for example, the research productivity of the student while in graduate school.

Our first alternative matching approach involves matching each Fulbright to a non-Fulbright who is the closest to them in terms of the predicted probability of being a Fulbright. Each of our Fulbrights is matched to the control student in the same broad scientific field with the closest predicted probability of being a Fulbright as based on a logit model of Fulbright status as a function of the exogenous variables previously used. 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

---

[35]These are: program rank, field dummies, Ph.D. year, GDP per capita of home country 5 years prior to Ph.D., gender and pregrad publications for specified regressions.

[36]An even more stripped-down specification would ignore all covariates and calculate the Wald estimator of the average difference in publications between pairs divided by the difference in the average share of years abroad, or its equivalent, 2SLS estimates similar to those above with no covariates. Not only does this ignore information about the timing of location and publication measures, but to estimate separate impacts of those from rich and poor countries, it requires us to drop all pairs with one person from a rich and the other from a poor sample, approximately 40% of the sample. As a result, we do not include the results in detail here. Not surprisingly, standard errors are quite large. Results give a consistently negative effect of the share of years spent abroad for low-income countries that are larger in absolute value than those for high-income sample, although only total and last-authored publications are significant at standard levels. For high-income countries, the effect of foreign location varies in sign and never approaches significance.

similar control to the Fulbright in question.[37] All covariates are balanced except the log of the GDP per capita of the home country five years before graduation.

The second alternative matching procedure approach is the Coarsened Exact Matching (CEM) procedure used by Azoulay et al. (2010) following Iacus, King, and Porro (2008). In this non-parametric procedure, one selects a set of covariates – in our case, 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 – and creates strata with an approximate (or "coarsened") match on these covariates. For each treated observation (where treatment in this case is Fulbright status), a matching control observation is drawn from within its stratum. The average control appears 1.319 times in the new dataset. Table 7 (right columns) shows the average values of the covariates among Fulbrights and controls matched using CEM. Again, only the log of the home country GDP per capita is not balanced.

For each of these matching methods, we run two sets of regressions (using the Mullahy estimation procedure). In the first, each of our eight research output variables is modeled as a function of LAGFORLOC and the log of the GDP per capita of the home country five years before graduation, with the latter variable included because it was significantly different for Fulbrights and controls even after matching (using either method). We do not include additional controls because of the insignificant differences between groups. In the second, we run the standard dependent variables on LAGFORLOC and the (lagged log) GDP per capita of the present country, both instrumented as in our Table 4 results. The results appear in Table 7. The only substantive difference from Table 4 are that the CEM estimates are significantly negative only at the 25th percentile of GDP (with the exception of last-authored pubs and high-impact cites).

D. Effects by field of study

It is possible that the impact of foreign location depends on the field. We therefore experimented with several different interaction terms between fields and foreign location in our total publication equations. One question we asked was whether the negative impact of foreign location was limited to fields dependent on large laboratories. We found that in fields without big labs, the negative impact of foreign location was concentrated in poor countries. However, foreign location negatively impacts last-authorship in large-lab fields in both poor *and* rich non-US countries. Since last authorship in science signifies the laboratory head, this suggests that people in large-lab fields are participating in others' labs (perhaps in the US). These results are tentative since with people in early

[37] 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). More details about both matching methods are included in the online Appendix B.

stages of their academic careers, very few are heads of large labs.

Another question we investigated was whether foreign location increased publications for those in agricultural and related fields (such as water resources and entomology) because many countries of origin are more rural than the U.S. and this might be counteracting larger negative effects in other fields. Dividing the impact of foreign location by whether the field was agricultural-related, we find that the significant negative effect is indeed in the non-agricultural fields but that this too was only true for poor countries. In poor countries, the impact of foreign location was insignificant in agricultural fields yet always positive. For richer countries, no effect of foreign location – whether in agricultural or non-agricultural fields -- was significant and signs differed. Here, also, results should be interpreted carefully since only 14 percent of observations are in the agricultural fields.

V. Conclusion

In this paper, we have examined whether newly-minted U.S. Ph.D.s of foreign origin maximize their post-Ph.D. contributions to science if they remain in the U.S. Others have documented that the high concentration of star scientists and excellent universities seems to have led to positive externalities for others' research productivity, suggesting that foreign-born recipients of U.S. doctorates who return to home countries will be less productive than they would have been in the U.S. Additional advantages of U.S. location include financial resources, access to top journals and conferences, a culture of research competitiveness, and business support of basic research.

To do this, this paper compares publication histories for a sample of U.S. Ph.D. recipients from abroad using exogenously-determined variation in post-Ph.D. location to identify the causal effect of location on research output. The results suggest that those who remain in the U.S. are at an advantage in terms of higher rates of publications, citations and familiarity with recent literature compared to those in countries with low GDP per capita. This is particularly true in terms of publishing highly-cited science in high impact journals. However, those in countries with high GDP per capita – especially those countries in the top decile of GDP per capita – are just as likely to publish, to be cited and to remain current as those remaining in the U.S.

One caveat to the conclusions in this paper is that we have sampled only foreign-born scientists who received their Ph.D. education in the U.S. These scientists are more likely than their compatriots to have links to the U.S. research world and therefore may be hurt less by their non-U.S. location. Conversely, their productivity may suffer more from being abroad due to the lack of a local scientific network built up during graduate studies. In our future research, we will study the networks of international collaboration and citation of our U.S.-educated sample of foreign-born, measuring how links to the U.S. and other countries develop and grow, depending not only on the scientists' location but also on their fields, employment sector and research abilities.

Overall, our findings suggest that a scientist exogenously located in a country high in the income distribution can expect to be as productive in research as he or she would be in the U.S. This finding is fairly surprising in light of the high degree of concentration of top scientists at U.S. universities. It may reflect the mitigating factors discussed in Section II above, particularly other countries' increasing commitment to excellent science and to higher levels of science education, as well as easier international collaboration made possible by technological developments and encouraged by global "brain circulation" of U.S.-educated Ph.D.'s returning to their home countries.

Our results have important implications both for the future of global knowledge creation and for the future of U.S.'s position in the international scientific hierarchy. As less wealthy countries' per capita GDP converges towards wealthier ones, their scientists' productivity will also converge towards U.S. levels. This potentially could vastly increase the number of scientists around the world doing leading edge scientific research. As a result, world-wide knowledge will expand at increasing rates. However, our results also imply that the U.S.'s position as the world center of science and innovation is at risk. In the past, the U.S. could count on agglomeration externalities in knowledge creation to ensure that it would maintain its dominant position in the scientific world. Our research suggests that these externalities are not sufficient to ensure that scientists will only be successful inside the U.S. The U.S. presently maintains its high position because many people enjoy living here, because many foreigners receive their Ph.D.'s here and most stay, and because there remains considerable U.S. government financial support for basic scientific research. If governments in Asia and Europe increase their efforts to lure scientists to their countries – particularly by offering large amounts of financial support – and continue to grow their own Ph.D. programs, our findings suggest that the U.S. could easily lose its dominance in scientific research. If this occurs, it will have a profound negative impact on U.S. economic growth, since a major source of U.S. economic comparative advantage has been due to the commercialization of new technologies stemming from the cutting edge research performed in U.S. universities (Mowery and Rosenberg 1989).

The U.S. will only remain at the forefront of science if it can continue to attract excellent S&E students to its graduate programs and if a large proportion of these graduates -- as well as significant numbers of other foreign scientists -- continue to choose to live in the U.S. This will require both increased government support of excellence in graduate science education and redoubled government efforts to fund scientific research.

References

Adams, James, Grant Black, Roger Clemmons, Paula Stephan (2005). “Scientific Teams and Institutional Collaborations: Evidence from U.S. Universities, 1981-1999.” *Research Policy,* 34(3), 259-85.

Agrawal, Ajay, Iain Cockburn, and John McHale (2006). “Gone but not forgotten: Knowledge Flows, Labor Mobility, and Enduring Social Relationships.” *Journal of Economic Geography* 6(5), 571-91.

Agrawal, Ajay, Devesh Kapur, and John McHale (2007). “Birds of a Feather - Better Together? Exploring the Optimal Spatial Distribution of Ethnic Inventors.” *NBER Working Paper 12823.*

Angrist, Joshua (2001). “Estimations of Limited Dependent Variable Models with Dummy Endogenous Regressors: Simple Strategies for Empirical Practice.” *Journal of Business and Economic Statistics*, 19(1), 2-16.

Angrist, Joshua , Guido Imbens and Donald Rubin (1996). “Identification of Causal Effects Using Instrumental Variables”, *Journal of the American Statistical Association,* 91(434), 444-455.

Angrist, Joshua and Guido Imbens (1995). “Two-Stage Least Squares Estimation of Average Casual Effects in Models with Variable Treatment Intensity.” *Journal of the American Statistical Association,* 90 (June), 431-442.

Audretsch, David and Maryann Feldman (1996). “R&D Spillovers and the Geography of Innovation and Production.” *American Economic Review*. 86(3), 630-40.

Audretsch, David, Erik Lehmann and Susanne Warning (2005). “University spillovers and new firm location.” *Research Policy* 34, 1113–1122.

Azoulay, Pierre, Joshua Graff-Zivin and Jialan Wang (2010), “Superstar Extinction”, *The Quarterly Journal of Economics* (2010) 125(2): 515-548

Bjorklund, Anders and Robert A. Moffitt (1987).“The Estimation of Wage Gains and Welfare Gains in Self-Selection Models.” *Review of Economics and Statistics*. 69(1).

Bound, John, Sarah Turner, and Patrick Walsh. *forthcoming*. “Internationalization of U.S. Doctorate Education.” In Richard Freeman and Daniel Goroff (eds.), *Brainpower, Science and Engineering Careers in the United States*. Chicago: University of Chicago Press.

Finn, M. G. (2007), “Stay Rates of Foreign Doctorate Recipients from U.S. Universities, 2005”, Working paper, Oak Ridge Institute for Science and Education.

Freeman, Richard (2006). “Does Globalization of the Scientific/Engineering Workforce Threaten U.S. Economic Leadership?” in Adam Jaffe, Josh Lerner and Scott Stern editors. *Innovation Policy and the Economy.* Cambridge MA: NBER/MIT Press.

Furman, Jeffrey, Michael Porter, and Scott Stern (2002), “The Determinants of National Innovative Capacity”, *Research Policy* 31, 899-933.

Goldberger, Marvin L., Brendan A. Maher, Pamela Ebert Flattau, Editors; (1995), *Research Doctorate Programs in the United States: Continuity and Change*, Committee for the Study of Research-Doctorate Programs in the United States. National Research Council. National Academy of Sciences: Washington, D.C.

Heckman, James J. and Edward Vytlacil (1999). “Local Instrumental Variables and Latent Variable Models for Identifying and Bounding Treatment Effects.” *Proceedings of the National Academy of Sciences*, 96(8).

Heckman, James J., Sergio Urzua, Edward Vytlacil (2006). “Understanding Instrumental Variables in Models with Essential Heterogeneity.” *Review of Economics and Statistics*, 88(3).

Iacus, SM, King G, Porro G. In Press. Causal Inference Without Balance Checking: Coarsened Exact Matching. *Political Analysis.*

Imbens, Guido and Angrist, Joshua (1994). Identification and Estimation of Local Average Treatment Effects. *Econometrica*, 62(2), 467-75.

Imbens, Guido and Jeffrey Wooldridge (2007). “What’s New in Econometrics? Lecture 5.” Cambridge MA: NBER Summer Workshop manuscript.

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.

Jaffe, Adam, Manuel Trajtenberg, Rebecca Henderson (1993). “Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations.” *Quarterly Journal of Economics*, 108:3, 577–98.

Jaffe, Adam (1989).“Real Effects of Academic Research.” *American Economic Review* 79(5), 957-70.

Kerr, William R. (2008). "Ethnic Scientific Communities and International Technology Diffusion." *Review of Economics and Statistics* 90 (3), 518-37.

Kim, E. Han, Adair Morse and Luigi Zingales (2006). Are Elite Universities Losing Their Competitive Edge? *NBER Working Paper 12245.*

Moffitt, Robert (forthcoming). Estimating Marginal Treatment Effects in Heterogeneous Populations.” *Annales d’Economie et de Statistique.*

Mowery, D. and N. Rosenberg (1989). *Technology and the Pursuit of Economic Growth*, Cambridge: Cambridge University Press.

Mullahy J. (1997) “Instrumental variable estimation of Poisson regression models: application to models of cigarette smoking behavior”, *Review of Economics and Statistics* 79, 586-593.

National Science Foundation, Division of Science Resources Statistics (2010). *Science and Engineering Indicators 2010.* Arlington, VA (NSB 10-01).

National Science Foundation, Division of Science Resources Statistics (2008). *Science and Engineering Indicators 2008.* Arlington, VA (NSB 08-01).

Orlando, Michael J. (2004) "Measuring spillovers from industrial R&D: on the importance of geographic and technological proximity." *Rand Journal of Economics*, 35(4), 777-786.

Oyer, Paul (2006). "Initial Labor Market Conditions and Long-Term Outcomes for Economists." *Journal of Economic Perspectives*, 20(3), 143.

Saxenian, Annalee (2002). "Brain Drain or Brain Circulation: How High-Skill Immigration Makes Everyone Better Off." *The Brookings Review*, Winter 20(1), 28-31.

Shanghai Jiao Tong University. *Academic Rankings of World Universities* 2003, 2007, and 2008. *www.arwu.org*.

Singh, Jasjit (2005). "Collaborative Networks as Determinants of Knowledge Diffusion Patterns" *Management Science*, 51(5), 756-770.

Stephan, Paula and Sharon Levin (1992). *Striking the Mother Lode in Science: The Importance of Age, Place, and Time.* New York and Oxford: Oxford University Press.

Stephan, Paula (1996) "The Economics of Science." *Journal of Economic Literature*, 34(3), 1199-1235.

Stock, James H. and Motohiro Yogo (2005) "Testing for Weak Instruments in IV Regression," in *Identification and Inference for Econometric Models: A Festschrift in Honor of Thomas Rothenberg*. Donald W.K. Andrews and James H. Stock, eds. Cambridge Univ. Press, 80–108.

Thompson, Peter; Fox-Kean, Melanie (2005). "Patent Citations and the Geography of Knowledge Spillovers: A Reassessment." *American Economic Review* 95(4), 461-464.

Wooldridge, Jeffery M. (2002). *Econometric analysis of cross section and panel data.* Cambridge MA: MIT Press.

Zucker, Lynne and Michael Darby (2007). Star Scientists, Innovation and Regional And National Immigration. *NBER Working Paper 13547.*

Zucker, Lynne and Michael Darby (2006). Movement of Star Scientists and Engineers and High-Tech Firm Entry. *NBER Working Papers 12172.*

Zucker, Lynne, Michael Darby and Marilynn Brewer (1998). "Intellectual Human Capital and the Birth of U.S. Biotechnology Enterprises." *American Economic Review*, 88(1), 290-306.

Zucker, Lynne, Michael Darby, Jonathan Furner, Robert Liu and Hongyan Ma (2007). "Minerva Unbound: Knowledge Stocks, Knowledge Flows and New Knowledge Production." *Research Policy* 36, 850-63.

Table 1: Research Output by Author's Current Location

	Full sample	In U.S., Europe, Asia, Can. Israel, Oceania	In U.S., Latin America, Africa or Middle East
	Total Publication Count		
In U.S.	1.234	1.234	1.234
Abroad	0.755	0.925	0.545
Total	0.972	1.111	0.993
Abroad minus In U.S.	**-0.479**	**-0.309**	**-0.683**
% difference	**-39%**	**-25%**	**-55%**
	Publications in High-impact Journals		
In U.S.	0.549	0.549	0.549
Abroad	0.290	0.352	0.215
Total	0.407	0.471	0.430
Abroad minus In U.S.	**-0.259**	**-0.177**	**-0.334**
% difference	**-47%**	**-36%**	**-61%**
Observations	4150	3120	2911

Table 2: Publications, (Forward) Citations and Median (Backwards) Citation Lag by Post-Ph.D. Location & Fulbright Status

Variable	Obs.	Mean	Std. Dev.	Max
Scientists located in USA				
Total publications count	1881	1.234	1.826	8
Total forward citations count	1881	18.240	39.602	174
First-authored publications	1881	0.493	0.887	4
First-authored fwd citations	1881	6.036	15.431	72
Last-authored publications	1881	0.324	1.053	21
Last-authored fwd citations	1881	1.979	6.753	34
High-impact publications	1881	0.549	1.136	5
Fwd citations to high-impact pubs	1881	10.448	27.904	125
Median citation lag	772	1.850	0.538	3.541
Scientists located outside USA				
Total publications count	2269	0.755	1.392	8
Total forwardd citations count	2269	7.906	25.397	174
First-authored publications	2269	0.322	0.787	12
First-authored fwd citations	2269	2.716	9.859	72
Last-authored publications	2269	0.240	0.758	10
Last-authored fwd citations	2269	1.410	5.633	34
High-impact publications	2269	0.290	0.779	5
Fwd citations to high-impact pubs	2269	4.528	18.116	125
Median citation lag	631	2.021	0.536	3.892
Control Scientists				
Total publications count	2108	1.058	1.673	8
Total forwardd citations count	2108	13.976	34.662	174
First-authored publications	2108	0.416	0.820	8
First-authored fwd citations	2108	4.327	12.907	72
Last-authored publications	2108	0.315	1.029	21
Last-authored fwd citations	2108	1.756	6.269	34
High-impact publications	2108	0.476	1.041	5
Fwd citations to high-impact pubs	2108	8.318	24.614	125
Median citation lag	733	1.908	0.556	3.892
Fulbright Scientists				
Total publications count	2042	0.883	1.560	8
Total forwardd citations count	2042	11.160	31.159	174
First-authored publications	2042	0.383	0.856	12
First-authored fwd citations	2042	4.111	12.685	72
Last-authored publications	2042	0.240	0.753	10
Last-authored fwd citations	2042	1.577	6.069	34
High-impact publications	2042	0.336	0.877	5
Fwd citations to high-impact pubs	2042	6.068	21.7160	125
Median citation lag	670	1.948	0.529	3.497

Table 3: Distribution of Controls and Fulbrights, by first-listed field of study

	Controls	Fulbrights	Total
Agricultural Sciences	30	34	64
Biological Sciences	47	53	100
Engineering & Computer Sciences	86	82	168
Earth/Air/Ocean Sciences	21	17	38
Mathematics & Statistics	21	22	43
Physical Sciences	27	23	50
Environment Science	12	13	25
Total	244	244	488

Table 4: Effect of Being Abroad on Publications and Citations

	(1) Total Publications	(2) First-authored Publications	(3) Last-authored Publications	(4) High-Impact Publications	(5) Total (fwd) Citations	(6) Cites to first-authored pubs	(7) Cites to last-authored pubs	(8) Cites to high impact pubs	(9) Median citation lag
A: Controls for home country GDP, no pregrad publications									
LAGFORLOC	-0.880**	-0.568	-0.964**	-1.420**	-0.706	+0.060	-0.588	-1.656**	0.160
	(0.359)	(0.378)	(0.435)	(0.557)	(0.436)	(0.527)	(0.584)	(0.679)	(0.123)
B: Controls for home country GDP and for pregrad publications									
LAGFORLOC	-0.612*	-0.311	-0.957**	-1.417**	-0.471	0.386	-0.815	-1.249*	0.159
	(0.346)	(0.381)	(0.464)	(0.593)	(0.439)	(0.644)	(0.641)	(0.643)	(0.125)
C: Percentage impact of Being Abroad by Percentiles of ln(per cap GDP current country), no controls for pregrad publications									
25th percentile	-0.848***	-0.778***	-0.894***	-0.932***	-0.946***	-0.806***	-0.919***	-0.989***	0.612**
50th percentile	-0.684***	-0.566***	-0.742***	-0.821***	-0.729***	-0.347	-0.676***	-0.891***	0.342**
75th percentile	-0.415**	-0.235	-0.452**	-0.594***	+0.0616	+0.825	+0.047	-0.210	0.113
90th percentile	-0.177	+0.049	-0.167	-0.360	+1.269	+2.231	+1.010	1.371	-0.015
D: Percentage impact of Being Abroad by Percentiles of ln(per cap GDP current country), with controls for pregrad publications									
25th percentile	-0.758***	-0.605***	-0.900***	-0.937***	-0.872***	-0.399	-0.933***	-0.966***	0.555**
50th percentile	-0.572***	-0.403*	-0.752***	-0.828***	-0.606***	+0.087	-0.743***	-0.833**	0.313*
75th percentile	-0.305	-0.155	-0.464**	-0.598***	+0.024	+0.796	-0.196	-0.351	0.108
90th percentile	-0.091	+0.026	-0.177	-0.355	+0.740	+1.373	+0.516	+0.382	-0.005
E: Subsample from rich home countries, controls for home country GDP, not pregrad publications									
LAGFORLOC	+0.256	+0.647	+2.063	+0.514	+0.477	+1.641*	+3.799	+0.944	-0.427
F. Subsample from rich home countries, controls for home country GDP and pregrad publications									
LAGFORLOC	+0.374	+0.543	+2.532	+1.400	+1.130	+2.755	+4.449**	+2.363**	-0.345
G. Subsample from poor home countries, controls for home country GDP, not pregrad publications									
LAGFORLOC	-1.005**	-0.444	-0.989**	-1.936***	-1.127**	+0.110	-1.232	-2.985***	0.284*
H: Subsample from poor home countries, controls for home country GDP and pregrad publications									
LAGFORLOC	-0.670*	-0.039	-1.048**	-1.878***	-1.022**	0.109	-1.762**	-2.604***	0.271*

* significant at 10%; ** significant at 5%; *** significant at 1% Robust standard errors, clustered by scientist. Given in parentheses for A and B.
Columns (1) – (8) estimated as Mullahy count-data IV model. Column (9) estimated as 2SLS.
Endogenous explanatory variables Panels A, B, E-H: Lagged foreign location (LAGFORLOC) Instrument: Fulbright dummy
Endogenous explanatory variables panels C and D: LAGFORLOC, LAGFORLOC*ln(lagged)current GDP. Instruments: Fulbright dummy and ln home country GDP 5 years prior to graduation.
Exogenous explanatory variables: scientific field dummies, period dummies, Ph.D. year dummies, gender, ln(rank of Ph.D. institution). In specified regressions, also include ln(real GDP per capita in home country 5 years prior to Ph.D. completion) and pregraduate total, first, and first+high+impact publications.

Table 5: Sensitivity Analysis of LIML Estimates of the Effect of Being Abroad by Income Level of Host Country

Percentile	(1) Total Publications	(2) First-authored Publications	(3) Last-authored Publications	(4) High-Impact Publications	(5) Total (fwd) Citations	(6) Cites to first-authored pubs	(7) Cites to last-authored pubs	(8) Cites to high impact pubs
ASSUMPTION: ZERO BIAS								
Average impact	-0.552**	-0.126	-0.260*	-0.412**	-9.412*	-1.018	-0.854	-7.252*
25th percentile	-1.780***	-0.476*	-0.967***	-1.009***	-35.756***	-5.295*	-4.728**	-23.484***
50th percentile	-1.040***	-0.262	-0.541**	-0.643***	-19.644***	-2.640	-2.361*	-13.513***
75th percentile	-0.414	-0.082	-0.180	-0.334**	-6.004	-0.392	-0.357	-5.072
90th percentile	-0.065	-0.019	+0.021	-0.163	+1.581	+0.857	+0.758	-0.378
ASSUMPTION: NEGATIVE BIAS OF 1 QUARTILE								
Average impact	-0.552	-0.039	-0.208	-0.325**	-6.592	-0.019	-0.507	-5.660
25th percntile	-1.572**	-0.389	-0.915***	-0.921***	-32.935***	-4.297	-4.381**	21.891***
50th percentile	-0.832**	-0.175	-0.489**	-0.556**	-16.824**	-1.642	-2.014	-11.921**
75th percentile	-0.206	+0.005	-0.128	-0.247	-3.183	0.606	-0.010	-3.479
90th percentile	0.143	+0.105	+0.073**	-0.075	+4.401	1.855	1.105	+1.214
ASSUMPTION: POSITIVE BIAS OF 1 QUARTILE								
Average impact	0.759***	-0.213*	-.312**	-0.499***	-12.232**	-2.016	-1.201	-8.844**
25th percentile	-1.988***	-0.563**	-1.019***	-1.096***	-38.576***	-6.294**	-5.075***	-25.076***
50th percentile	-1.248***	-0.349***	-0.592***	-0.731***	-22.464***	-3.639*	-2.708**	-15.105***
75th percentile	-0.622**	-0.169	-0.232*	-0.422***	-8.824*	-1.391	-0.704	-6.664*
90th percentile	-0.273	-0.068	-0.031	-0.250*	-1.239	-0.141	+0.411	-1.970

* significant at 10%; ** significant at 5%; *** significant at 1%
Exogenous explanatory variables: scientific field dummies, period dummies, Ph.D.year dummies, gender, ln(rank of Ph.D. institution).
Endogenous explanatory variables: LAGFORLOC, LAGFORLOC X ln (lagged) current GDP.
Instruments: Fulbright dummy and ln home country GDP 5 yrs prior to graduation.

Table 6: Effect of Being Abroad on Publications and Citations
Cross-sectional data collapsed to scientist-level; Dependent variables averaged over years since Ph.D.

	(1) Total Publications	(2) First-authored Publications	(3) Last-authored Publications	(4) High-Impact Publications	(5) Total (fwd) Citations	(6) Cites to first-authored pubs	(7) Cites to last-authored pubs	(8) Cites to high impact pubs
A: Subsample from rich countries, controls for home country GDP, not pregrad publications								
% of years abroad	0.162	0.508	1.704	0.516	1.020	1.628*	2.932**	1.366
	(0.821)	(0.919)	(1.171)	(0.895)	(0.829)	(0.936)	(1.339)	(0.909)
B: Subsample from rich countries, controls for home country GDP and for pregrad publications								
% of years abroad	0.097	0.083	1.829	1.500	1.566*	2.374**	3.259***	3.055**
	(0.824)	(0.991)	(1.127)	(1.130)	(0.947)	(1.045)	(1.246)	(1.472)
C: Subsample from poor countries, controls for home country GDP, not pregrad publications								
% of years abroad	-1.021***	-0.673*	-1.073*	-1.743***	-0.801	-0.192	-1.986**	-1.943***
	(0.374)	(0.346)	(0.552)	(0.460)	(0.605)	(0.563)	(0.964)	(0.562)
D: Subsample from poor countries, controls for home country GDP and for pregrad publications								
% of years abroad	-0.736**	-0.259	-1.387**	-1.757***	-0.709	0.169	-2.477***	-1.782***
	(0.329)	(0.313)	(0.567)	(0.447)	(0.478)	(0.546)	(0.868)	(0.542)
E: Subsample from high article per capita countries, controls for home country GDP, not pregrad publications								
% of years abroad	0.496	1.202	2.439	0.967	0.809	2.060	3.571	1.270
	(0.902)	(1.204)	(1.860)	(1.237)	(1.176)	(1.703)	(2.223)	(1.534)
F: Subsample from high article per capita countries, controls for home country GDP and for pregrad publications								
% of years abroad	0.482	1.077	1.830	1.892	1.835	3.256*	4.009*	3.855*
	(0.821)	(1.384)	(1.419)	(1.169)	(1.143)	(1.777)	(2.105)	(2.233)
G: Subsample from low article per capita countries, controls for home country GDP, not pregrad publications								
% of years abroad	-1.072**	-0.557	-1.191*	-1.585***	-0.408	0.243	-1.674*	-1.634***
	(0.475)	(0.395)	(0.614)	(0.500)	(0.591)	(0.538)	(0.990)	(0.545)
H: Subsample from low article per capita countries, controls for home country GDP and for pregrad publications								
% of years abroad	-0.925**	-0.404	-1.608**	-1.864***	-0.557	0.206	-2.496***	-1.673***
	(0.424)	(0.371)	(0.649)	(0.528)	(0.480)	(0.527)	(0.935)	(0.557)

* significant at 10%; ** significant at 5%; *** significant at 1%Robust standard errors in parentheses.
Estimated with Mullahy model. Endogenous explanatory variable % of years abroad. Instrument: Fulbright dummy.
Exogenous explanatory variables: scientific field and PhD year dummies, gender, ln(rank of Ph.D. institution) and GDP per capita of home country 5 years prior to graduation. In specified regressions, also include pregraduate total, first, and first+high+impact publications.

Table 7
Estimation Using Alternative Matching Methods

	(1) Total Publications	(2) First-authored Publications	(3) Last-authored Publications	(4) High-Impact Publications	(5) Total (fwd) Citations	(6) Cites to first-authored pubs	(7) Cites to last-authored pubs	(8) Cites to high impact pubs
Panel A: Matching on Pr(Fulbright)								
LAGFORLOC	-0.231	-0.31	-0.311	-0.899	-0.384	0.412	0.027	-0.597
	(0.443)	(0.561)	(0.604)	(0.760)	(0.681)	(1.121)	(0.752)	(0.867)
Panel B: Matching on Pr(Fulbright), at different levels of GDP								
25th percentile	-0.856***	-0.852***	-0.926***	-0.903***	-0.945***	-0.862***	-0.919***	-0.974***
50th percentile	-0.563***	-0.573**	-0.721***	-0.757**	-0.628**	-0.127	-0.595	-0.865***
75th percentile	0.121	0.048	-0.143	-0.473	0.889	3.178	0.589	-0.462
90th percentile	0.893	0.726	0.601	-0.188	3.663	8.974	2.396	0.162
Panel C: CEM Matching								
LAGFORLOC	-0.122	0.753	-0.519	-0.562	-0.131	2.175	-0.298	-0.524
	(0.430)	(0.614)	(0.634)	(0.617)	(0.576)	(3.019)	(0.767)	(0.694)
Panel D: CEM Matching, at different levels of GDP								
25th percentile	-0.694***	0.113	-0.851***	-0.817***	-0.883***	282816	-0.841***	-0.939***
50th percentile	-0.345	0.824	-0.557***	-0.488	-0.067	6533.629	-0.326	-0.616**
75th percentile	0.245	1.771	0.11	0.221	4.43	268.058	1.29	0.816
90th percentile	0.78	2.497	0.852	0.979	13.454	44.662	3.521	3.304

* significant at 10%; ** significant at 5%; *** significant at 1%

Appendix Table: First stage regressions

	(1) Full sample	(2) Full sample, including interaction with GDP	(3)	(4) Subset: Rich home country	(5) Subset: Poor home country
dependent variable:	LAGFORLOC	LAGFORLOC	LAGFORLOC X ln GDP	LAGFORLOC	LAGFORLOC
Fulbright dummy	0.402 (0.036)***	0.406 (0.036)***	3.708 (0.339)***	0.184 (0.063)***	0.475 (0.045)***
Real gdp per cap of home country 5 yrs prior to grad	0.062 (0.023)***			-0.381 (0.156)**	0.102 (0.032)***
Biological sciences	-0.22 (0.059)***	-0.224 (0.059)***	-1.97 (0.552)***	-0.189 (0.102)*	-0.248 (0.070)***
Engineering & Computer Science	-0.161 (0.056)***	-0.165 (0.055)***	-1.346 (0.519)***	-0.172 (0.099)*	-0.162 (0.062)***
Earth/Air/Ocean Sciences	-0.045 -0.088	-0.049 -0.088	-0.208 -0.843	0.025 -0.155	-0.074 -0.109
Mathematics & Statistics	-0.078 -0.078	-0.079 -0.078	-0.571 -0.73	0.049 -0.118	-0.139 -0.1
Physical Sciences	-0.208 (0.065)***	-0.213 (0.065)***	-1.803 (0.625)***	-0.162 -0.1	-0.174 (0.081)**
Environmental Science	-0.048 -0.087	-0.05 -0.086	-0.237 -0.811	0.107 -0.158	-0.068 -0.098
Year of PhD = 9596	0.206 (0.075)***	0.192 (0.075)**	1.831 (0.718)**	0.166 -0.139	0.163 (0.093)*
Year of PhD = 9798	0.146 (0.072)**	0.138 (0.072)*	1.252 (0.691)*	-0.001 -0.126	0.179 (0.091)**
Year of PhD = 9899	0.045 -0.074	0.042 -0.074	0.316 -0.71	-0.097 -0.12	0.097 -0.093
Year of PhD = post02	0.126 -0.116	0.125 -0.116	1.03 -1.093	0.046 -0.292	0.143 -0.131
Year of PhD = pre95	0.424 (0.083)***	0.404 (0.084)***	3.788 (0.789)***	0.238 -0.155	0.458 (0.111)***
Year =9798	-0.068 (0.027)**	-0.059 (0.027)**	-0.625 (0.259)**	-0.075 -0.046	-0.044 -0.031
Year =9899	-0.018 -0.015	-0.013 -0.015	-0.128 -0.143	0.001 -0.024	-0.026 -0.018
Year =post02	0.069 (0.014)***	0.06 (0.015)***	0.59 (0.139)***	0.056 (0.024)**	0.072 (0.017)***
Year =pre97	-0.105 (0.049)**	-0.093 (0.049)*	-0.996 (0.459)**	-0.183 (0.087)**	-0.05 -0.05
ln(Rank of Ph.D program)	-0.002 -0.016	-0.002 -0.016	-0.059 -0.151	-0.047 (0.026)*	0.001 -0.02
Female	0.024 -0.044	0.025 -0.044	0.229 -0.423	0.03 -0.072	0.049 -0.054
lagged GDP per capita of home country		0.059 (0.023)**	0.878 (0.219)***		
Constant	-0.26 -0.238	-0.241 -0.244	-5.229 (2.298)**	4.327 (1.610)***	-0.641 (0.299)**
Observations	4150	4150	4150	1419	2731
R-squared	0.26	0.26	0.27	0.24	0.33

Online Appendix A: Data Appendix
Note to Editor: Will be made available on author's website instead of published

Fulbright Data

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.

Location Search Procedure

First, we 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 the student's post-Ph.D. locations and affiliations. The search time was limited 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. Click on the name of the student's advisor. If this step failed (i.e. there are 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.

For schools listing prior degrees or biographical information in the dissertation, we used this information to infer the student's country of origin (see below). For schools that did not list prior degrees, if we found a potential control student, we looked them up on the web. If we could find their current location and evidence that they came from a foreign country (i.e. foreign undergraduate degree or biography), we recorded their name, year of Ph.D., current location, and estimated country of origin.

ProQuest Dissertations and Theses

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.

Several universities require students to list biographical information in the front matter of the dissertation. Table A1 lists these universities, which were identified by checking dissertations filed at the universities that are major producers of scientists and engineers in the United States. 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 A1 and A2 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, the *SED* lists the country of undergraduate degree. For 84.9% of students, the country of undergraduate degree is the same as the country of citizenship. However, there is considerable heterogeneity across countries in the extent to which students pursue undergraduate studies outside their countries of origin. Table A2 presents, for a selected list of countries, the share of students responding to the *SED*'s questions who remained in their home country for undergraduate study. Students from Germany and Japan have the lowest rates of staying at home among the major producers of U.S. graduate students (73% and 74%, respectively). However, the countries that send the most students (China, India, Taiwan, Korea, and Canada) have high stay-at-home rates for undergraduate study (98%, 93%, 89%, 76%, and 82%, respectively). Furthermore, 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.*

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, 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, complete backward citations, 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 median backward citation lag also includes self-citations. In future work, we intend to remove these citations. However, this requires downloading bibliographic data on each specific citing article, which is a very time-consuming process.

The *ISI Web of Science* database does not cover every scientific journal published worldwide. It lists articles from 6,650 scientific journals. 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 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 (0.36% of our observations are on scientists located in China, 0.55% in Japan, 0.87% in Korea, 1.03% in Taiwan and 1.5% in Thailand). 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).[38] 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.

Match Statistics

For exactly 122, or exactly 50%, of pairs, the Fulbright and control graduated in the same year, and 76.7% 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).

For 33% of the pairs, the control and Fulbright come from the same region of origin. Europeans paired with other Europeans make up 22% of the sample. Among those with discordant regions of origin, the most common pattern was Latin American Fulbrights matched with Asian controls (representing 18% of the sample). The second most common pairing was Latin American

Fulbrights with European controls (17% of the sample), followed by European Fulbrights paired with Asian controls (8% of the sample).

There are 67 pairs, or 27% of the sample, in which the advisor is the same for both members. The broadly-defined field is the same for 83% 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.

Figure A3 and Tables A3 and A4 describe differences between Fulbrights and controls in terms of region, country, and Ph.D. year.

Table A1: Universities listing biographic info in thesis

AUBURN
BOSTON U
CALIFORNIA STATE U
CLARK
CORNELL U
FLORIDA INSTITUTE OF TECHNOLOGY
FORDHAM
GEORGE WASHINGTON U
GEORGETOWN U
KANSAS STATE
LOUISIANA STATE U
NC STATE
OH STATE
OK STATE
SYRACUSE
TEXAS A&M
U ARKANSAS
U CALIFORNIA
U CINCINATTI
U COLORADO
U CONNECTICUT
U FLORIDA
U ILLINOIS
U MAINE
U MASSACHUSETTS
U MASSACHUSETTS AT AMHERST
U MISSOURI
U NEVADA
U OREGON
U PITTSBURGH
U SOUTH ALABAMA
U SOUTH CAROLINA
U VIRGINIA

Table A2: Share of Ph.D. students at U.S. universities who received undergraduate degrees in their countries of citizenship

AUSTRALIA	85.00%
BRAZIL	96.02%
CANADA	82.51%
CHINA	98.35%
EGYPT	96.38%
FRANCE	82.05%
GERMANY	73.05%
GREECE	80.51%
INDIA	92.71%
IRAN	88.33%
ISRAEL	88.46%
JAPAN	73.51%
MEXICO	89.19%
NIGERIA	60.61%
PHILIPPINES	87.23%
SOUTH KOREA	76.33%
TAIWAN	89.19%
THAILAND	87.28%
TURKEY	95.57%
U.K.	63.64%
Weighted average across these countries	89.50%
Weighted average across all countries	84.79%

Table A3: 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	93	102
Brazil	11	0	11	Morocco	0	2	2
Bulgaria	1	0	1	Netherlands	4	5	9
Canada	8	0	8	Nigeria	2	0	2
Chile	3	0	3	Norway	2	6	8
China	18	0	18	Pakistan	2	0	2
Colombia	4	8	12	Panama	1	1	2
Costa Rica	0	3	3	Peru	2	2	4
Cote D'Ivoire	0	2	2	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	5	1	6
Denmark	2	4	6	Russia	9	0	9
Ecuador	1	0	1	Singapore	1	0	1
Egypt	2	0	2	Solomon Islands	0	1	1
Ethiopia	2	2	4	South Africa	0	7	7
Finland	2	5	7	Spain	6	7	13
France	2	0	2	Sri Lanka	1	0	1
Germany	10	0	10	Swaziland	1	0	1
Ghana	0	2	2	Sweden	2	3	5
Greece	4	7	11	Switzerland	3	1	4
Guatemala	1	1	2	Taiwan	7	0	7
Haiti	0	1	1	Tanzania	1	1	2
Hungary	3	1	4	Thailand	5	5	10
Iceland	2	7	9	Togo	0	2	2
India	25	0	25	Trinidad & Tobago	1	1	2
Indonesia	4	0	4	Turkey	11	1	12
Iran	1	0	1	UK	2	4	6
Iraq	1	0	1	Uganda	1	2	3
Ireland	2	1	3	Ukraine	5	0	5
Israel	3	6	9	Venezuela	2	1	3
Italy	5	3	8	Yugoslavia	3	0	3
Japan	5	0	5	Zimbabwe	1	0	1
Jordan	1	0	1	Total	244	244	488

Table A4: 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	7	5	12
1994	15	17	32
1995	11	23	34
1996	31	27	58
1997	45	36	81
1998	38	40	78
1999	33	34	67
2000	28	22	50
2001	13	22	35
2002	9	10	19
2003	7	6	13
2004	2	1	3
2005	2	1	3
Total	244	244	488
Average	1997.881	1997.897	1997.889

Figure A1

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 A2

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

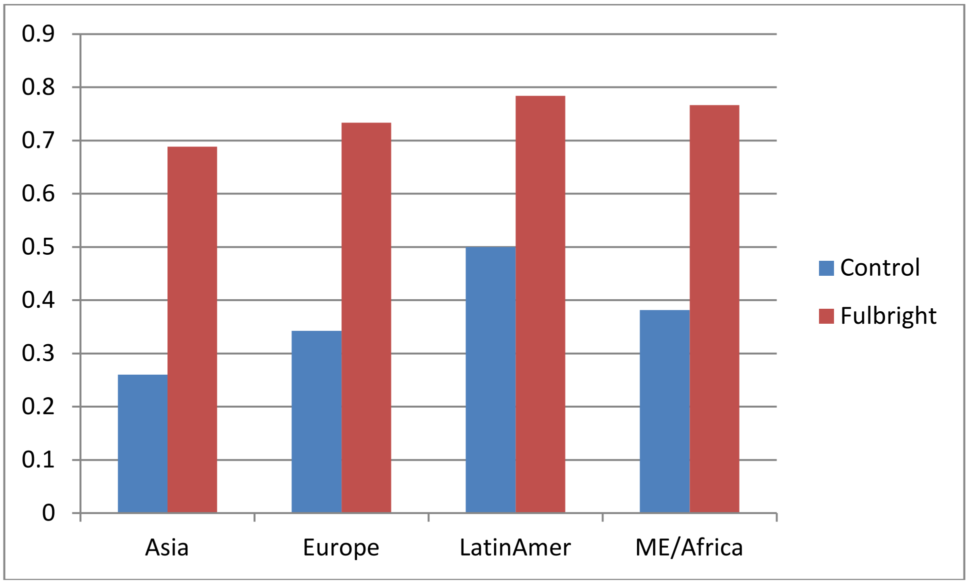
P. Zhang, and M. Ismail "A New RF Front-End and Frequency Synthesizer Architecture for 3.1–10.6 GHz MB-OFDM UWB Receivers", *Proc.* 48^{th} *Midwest Symposium on Circuit and System*, vol.2, pp.1119–1122, August 2005.

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.* 48^{th} *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.* 48^{th} *Midwest Symposium on Circuit and System*, vol.1, pp.187–190, August 2005.

vii

Figure A3: % of post-Ph.D. years spent abroad, by region of origin



Note: "Europe" includes Australia, Canada, and Israel.

Online Appendix B: Details on Alternative Matching Methods

Our first alternative matching approach involves matching each Fulbright to a non-Fulbright who is the closest to them in terms of the predicted probability of being a Fulbright.[39] We estimate a logit model in which the dependent variable is the Fulbright dummy and explanatory variables are the exogenous variables previously used. The pseudo-R^2 of this regression was 0.218. 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.[40] 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 205 Fulbrights and 77 controls in the final dataset, with the typical control appearing 2.6 times in the re-matched dataset. The average program rank is for controls 32.5 for controls and 34.3 for Fulbrights. 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 except the log of the GDP per capita of the home country five years before graduation.

The second alternative matching procedure approach is the Coarsened Exact Matching (CEM) procedure used by Azoulay et al. (2010) following Iacus, King, and Porro (2008). In this non-parametric procedure, the analyst selects a set of covariates on which to achieve balance between treated and control observations. A set of strata are then created to cover the support of the joint distribution of the covariates, and for each treated observation (where treatment in this case is Fulbright status), a matching control observation is drawn from within its stratum.

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 power. Therefore, we "coarsen" the distribution of all covariates but the field of study, on which we require an exact match.[41]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 488 controls and Fulbrights, we must use fairly coarse bins to avoid

[39] 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).

[40] 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.

[41]Note that field is already considerably coarsened, as we have grouped together narrower fields into broad subject areas.

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 388 individuals while dropping 100 (22 Fulbrights and 78 controls) for whom a match was not found.[42] The average control appears 1.319 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.

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

	matching on P(Fulbright)			**CEM**		
	Mean for Controls	Mean for Fulbrights	p-value of t-test of difference in means	Mean for Controls	Mean for Fulbrights	p-value of t-test of difference in means
# Observations	77	205		166	222	
LAGFORLOC	0.356	0.756	0.000	0.351	0.775	0.000
Real GDP per cap of hc 5 yrs prior to PhD	9.186	8.963	0.043	8.703	8.911	0.021
Pre-PhD 1st authored pubs	1.649	1.341	0.237	1.187	1.252	0.723
Log(university rank)	3.086	3.000	0.574	3.074	3.132	0.628
Year of PhD	1997.675	1997.888	0.514	1997.837	1997.896	0.816

[42] If, instead, we coarsen each covariate into terciles, we end up with only 179 individuals in the dataset.