Best and Brightest? The Selectivity of Foreign-Born Ph.D. Recipients in the U.S.*

Valerie Bostwick, Joseph Staudt, Bruce A. Weinberg †

January 2023

PRELIMINARY: PLEASE DO NOT CITE OR CIRCULATE.

Abstract

The foreign-born are central to the STEM workforce in the U.S. This paper examines the selectivity of U.S.-trained, foreign-born Ph.D. recipients relative to their U.S.-born and trained counterparts. We do this in terms of both the training/laboratory environments to which they were exposed during graduate school as well as their post-graduation labor market outcomes. We uncover strong evidence of selection, both pre- and post-graduation, and across the observed choices Ph.D. recipients make (e.g. field or sector of employment) as well as their unobserved characteristics (e.g. "ability"). Moreover, we find that as the strength of Ph.D. recipients' attachment to the U.S. labor market increases, so does the selectivity of the foreign-born. Thus, at least among Ph.D. recipients, the world does not appear to be sending the United States their tired and poor, but rather their best and brightest, yearning to succeed.

JEL Classification Numbers:
Keywords:

*Thanks ... All errors are ours. Any opinions and conclusions expressed herein are those of the authors and do not represent the views of the U.S. Census Bureau. The Census Bureau has ensured appropriate access and use of confidential data and has reviewed these results for disclosure avoidance protection (Project 7507193: CBDRB-FY24-CES022-001).

†Bostwick: Kansas State University and IZA, vkbostwick@ksu.edu; Staudt, U.S. Census Bureau, joseph.staudt@census.gov; Bruce A. Weinberg, Ohio State University, IZA, and NBER, weinberg.27@osu.edu.

1 Introduction

Immigrants comprise nearly half of U.S. STEM workers with doctoral degrees (Kerr and Lincoln, 2010) and contribute disproportionately to U.S. innovation and long-run growth.[1] As the United States strives to maintain its global competitiveness and leadership in innovation, substantial public funds are directed toward nurturing, at least implicitly, the training of foreign-born Ph.D. students. Given the stakes for U.S. economic performance, the share of STEM researchers who are foreign-born, and the public investment supporting foreign-born trainees, it is important to assess whether the U.S. is attracting and retaining the "best and brightest" or a less desirable slice of STEM trainees. It is also important to assess whether the people we attract are in fields that are in demand and whether they are crowding out domestic trainees.

This paper examines the selectivity of U.S.-trained, foreign-born Ph.D. recipients supported on research grants, benchmarking them against their U.S.-born and trained counterparts. Our unique data allow us to make this comparison along many dimensions, including the research / laboratory environments that trainees experienced during graduate school as well as their post-graduation labor market outcomes. Overall, there is strong evidence of positive selection, both pre- and post-graduation, and across the observed choices Ph.D. recipients make (e.g. field or sector of employment) as well as their unobserved characteristics (e.g. "ability"). Moreover, as the strength of Ph.D. recipients' attachment to the U.S. labor market increases, so does the selectivity of the foreign-born. Thus, at least among Ph.D. recipients, the U.S. appears to be attracting and retaining the "best and brightest" in economically critical fields.

The data that makes our analysis possible is a unique match between micro-level Census data on individual characteristics, labor market outcomes, and employer characteristics and data from the UMETRICS program, which provides information on research environments for trainees (and others) supported on sponsored projects. These measures include funding duration, timing, and source. These data are, in turn, linked to ProQuest's Dissertation and Theses data, which allows us to identify each student's degree year and field of specialization. This linked dataset provides several advantages over previous studies, which have largely relied on the Survey of Earned Doctorates (SED). For instance, We are able to observe administrative records on employment and earnings in the U.S., whereas the SED reports the intention to stay in the U.S. and has many gaps in reporting. We follow graduates for 4 years post-degree, tracking whether a graduate is working in academia or industry and how

[1] Hunt and Gauthier-Loiselle (2010) shows that increases in the share of foreign-born college graduates leads to increases in the number of patents per capita.

earnings and sector change over time. Moreover, we can control for several aspects of the research environment in which people trained.

We begin with an analysis of the fields that people study. We show that there are stark differences in the fields in which foreign- and native-born Ph.D.s study. Foreign-born students are nearly twice as likely to study engineering as native-born students (and slightly more likely to study mathematics). By contrast, native-born students are more likely to study biology and health / medicine. These differences are highly consequential, with mathematics and engineering being the two highest paying fields and biology and health / medicine being among the lowest. This positive relationship between earnings in a field and the representation of foreign-born relative to native-born is quantitatively meaningful - our analyses show that field of study accounts for 61.6% of the earnings gap between foreign- and native-born Ph.D. recipients. It is also important qualitatively in that it indicates that the foreign-born are studying in fields that are in high demand. It also suggests that crowding out of native-born students is likely small, as the market can bear a greater quantity supplied in these fields.

We then turn to the training environment of U.S.- and foreign-born Ph.D. recipients and compare the research settings they experience while in graduate school. The U.S.-born tend to work in labs with larger teams, a lower faculty to other personnel ratio, are paid by fewer awards, and are paid by larger awards. To assess the overall value of the various characteristics of the training environment, we regress earnings on characteristics and use the coefficients to generate a composite measure for the quality of the training environment. We find that, on average, foreign-born workers experience better training environments, but that these differences explaining a very small portion (2-3%) of the log earnings gap between U.S.- and foreign-born Ph.D. recipients.

Next, we examine post-graduation earnings differences between foreign-born and U.S.-born Ph.D. recipients. Our analysis shows a 17% wage premium for the foreign-born that, though attenuated, remains at 4.5% after controlling for various factors including sector of employment, field, age, gender, Ph.D.-granting institution, and the characteristics of the lab in which they worked while in graduate school. Nearly all the attenuation is due to controls for sector of employment and field. Thus, as indicated, the foreign born are much more likely to receive their their Ph.D. in fields with more lucrative career outcomes, such as engineering, as opposed to lower paying fields such as biology. The foreign-born are also more likely than their U.S.-born peers to take a higher-paying industry job rather than an academic job.

The earnings premium experienced by the foreign-born persists even after controlling for our rich set of individual-level characteristics, suggesting that the foreign-born may be further

positively selected (or willing to work longer hours and/or in jobs that offer compensating differences). Moreover, we provide evidence of continued selection of the best foreign-born graduates into the U.S. labor force after initial job placements. The foreign-born earnings premium is growing over time from degree. After 4 years of post-degree employment, foreign-born doctoral recipients are earn 6.2% more than their U.S.-born counterparts. Furthermore, the sample of foreign-born workers who are still employed in the U.S. after 4 years have higher initial earnings than their peers who eventually leave the U.S. labor market. All told, or results strongly suggest that we retain the "best and brightest" foreign-born students, even within field.

Our results are consistent with previous research by Grogger and Hanson (2015), who find that foreign-born STEM Ph.D.s with the strongest initial academic potential (as measured by their attributes at the time of entering graduate school) are the most intent on staying in the U.S. after graduation. We build on this, showing that those who stay are positively selected and earn higher earnings – even within field and sector of employment. We also document that this is not a temporary phenomenon by documenting the longer-term stay rates and earnings premia of foreign-born STEM Ph.D.s 4 years after graduation.

These findings have important implications for understanding the dynamics of the labor market and the role of immigration in shaping it as well as for policymakers interested in retaining the best foreign students to bolster the US economy.

2 Literature Review

The best estimates of the stay rate for foreign students receiving Ph.D.s comes from the NSF's Survey of Doctorate Recipients. The 5-yr stay rate for all foreign students receiving doctorates in the years 2011-2013 was 71%.[2] Grogger and Hanson (2015) use the Survey of Earned Doctorates from 1960-2008 to show that there is positive selection in the foreign-born STEM doctoral students who report that they are intending to stay in the US after graduation. This is measured indirectly through indicators such as parental education and having received merit-based financial support.

What role do preferences play in foreign-born Ph.D. stay rates? In a survey study focused specifically on U.S. STEM doctoral students from China, Zeithammer and Kellogg (2013) find that approximately 70% of Chinese STEM PhDs would prefer to return to China if offered the same salary as they expect to receive in the US, but that salary differentials between the two countries keep the majority of them in the US. By contrast, Ganguli and Gaulé (2019) conduct an original survey of doctoral students in chemistry and find that international

[2]https://ncses.nsf.gov/pubs/nsb20198/immigration-and-the-s-e-workforce See Table 3-22

students have stronger preferences for remaining in the U.S. for post-doctoral training than domestic students, while US students value teaching more. Another survey of STEM PhD students shows that, despite having stronger preferences for working in entrepreneurial firms, foreign-born Ph.D. recipients are less likely to be employed by such firms, possibly due to visa restrictions (Roach et al., 2019).

In a related literature, researchers have estimated the impacts of U.S. visa laws on foreign-born Ph.D.s' labor market decisions. Among foreign-born Ph.D.s who stay in US, permanent visa holders are 24% less likely to take postdoctoral positions than temporary visa holders (Lan, 2012). Amuedo-Dorantes and Furtado (2019) similarly find that H1-B visa caps (which don't apply to academic institutions) led to an increase in the probability of academic employment for foreign-born Ph.D.s. These binding visa caps raised international students' likelihood of employment in academia even outside of their field of study, which the authors argue is consistent with the notion of "settling" for academia. Khosla (2018) and Kahn and MacGarvie (2020) find that increases in wait time for permanent residency status for doctoral recipients from China and India are significantly associated with declines in U.S. stay rates.

There is some evidence that characteristics of the home country determine the stay rates of foreign-born Ph.D.s. Kahn and MacGarvie (2020) find that the strength of the science enterprise in the home country is an important predictor of stay rates, with students from the lowest-ranked countries in terms of publication output exhibiting stay rates that are 10pp higher than those from the highest-ranked countries. Research quality (as measured by citations per article) is also a strong predictor of return to the home country. Grogger and Hanson (2015) find that foreign-born STEM Ph.D.s tend to stay in the U.S. during periods of strong economic growth and during periods of weak home country growth.

3 Data

Our data are unique in that they allow us to identify the place of birth for a large sample of Ph.D. recipients who can also be linked to comprehensive tax records as well as a variety of detailed information on their demographics, degree, and the characteristics of the labs in which they worked while in graduate school. These linkages allow us to estimate the foreign-born earnings premium and examine the extent to which it is driven by observed choices (such as field or sector of employment) and unobservable characteristics (such as ability). They also allow us to examine whether the foreign- and US-born select into different training environments while in graduate school and assess whether these environments play a role in determining the foreign-born earnings premium.

3.1 UMETRICS

Our sample of graduate students is obtained from the UMETRICS, which is run by the Institute for Research on Innovation and Science (IRIS) at the University of Michigan (UMETRICS, 2018, Q4; Nicholls et al., 2022).[3] The data are derived from university payroll and human resources records and track all payments made by university grants to employees and vendors (Lane et al., 2015). Our vintage of UMETRICS includes grant transactions at XX major research universities, which account for more than one-third of federally funded academic R&D.

Graduate Students UMETRICS also includes information on the job titles of individuals receiving payments from grants, allowing us to identify their occupations, which include "faculty", "graduate student", "postdoc", "undergraduate", and "staff" (Ikudo et al., 2019). The UMETRICS graduate students comprise the core of our sample. Though UMETRICS graduate students are not representative of all graduate students, they are likely a fair representation of grant-supported graduate students in STEM fields at large research universities.

Training Environment A unique strength of UMETRICS data is that we can leverage its transaction structure to construct a variety of detailed measures of the characteristics of the labs in which graduate students worked while they were in graduate school – both in terms of funding and personnel. No other data source offers this opportunity at scale. We characterize a graduate student's lab environment using the following four variables: 1) the number of grants that pay the graduate student, 2) the average size (in dollars) of the grants that pay the graduate student, 3) the number of other personnel paid on the grants that pay the graduate student (team size), and 4) the ratio of faculty to other lab personnel on the grants that pay the graduate student. Using these lab covariates, we can examine differences in the typical training environment experienced by foreign-born and US-born Ph.D. recipients as well as determine whether these differences translate in to earnings gaps after graduation.

3.2 ProQuest Dissertations

IRIS provides a crosswalk between the ProQuest Dissertation and Theses Database (ProQuest, 2018) and UMETRICS employees, which we use to derive the subset of graduate

[3]The UMETRICS data are stored in a data enclave at IRIS. Annually, a snapshot of the data is added to the Federal Statistical Research Data Centers (FSRDCs).

students that can be linked to their dissertation.[4] Since not all UMETRICS graduate students will go on to earn their Ph.D., this linkage allows us to focus on the approximately 14,900 who successfully complete their graduate training. We can then determine these Ph.D. recipients' field of study, which we use to construct field fixed effects that are used in regressions. We can also determine the year in which each of these graduate students received their Ph.D., allowing us to track *post-degree* labor maket outcomes as well as construct degree-year fixed effects that are included in regressions.

3.3 Demographic Characteristics

We obtain demographic information on each of the Ph.D. recipients in our sample by linking them to the Individual Characteristics File (ICF), which is part of the infrastructure of the Longitudinal Employer-Household Dynamics (LEHD) program at Census (Vilhuber et al., 2014), and is derived from administrative and survey data from the Social Security Administration (SSA) and the U.S. Census Bureau.

Place of Birth For this paper, the most important piece of information we take from the ICF is each UMETRICS Ph.D. recipient's place of birth, which can be one of 35 region/country codes.[5] We code each Ph.D. recipient who is born in the United States as "US-born" and code all other Ph.D. recipients as "foreign-born".[6] It it worth noting that foreign-born status is not synonymous with citizenship. Indeed, a person can be born abroad and be a US-citizen or be born in the US but be a non-citizen. However, there is likely substantial overlap and place of birth is often used as a reasonable proxy for citizenship (Brown et al., 2018).

Other Demographic Characteristics In addition to a Ph.D. recipient's place of birth, we also use the ICF to obtain other demographic characteristics. In particular, we get each Ph.D. recipient's gender and age, which are used as covariates in regressions.

[4]Some graduate students link to multiple dissertations and some dissertations are linked to multiple graduate students. We only retain the graduate students that match, 1-to-1, to a ProQuest dissertation, which enables us to think of graduate students and dissertations interchangeably.

[5]Central Asia, South East Asia, Middle East and North Africa, Caribbean, Central America, South America , Africa, Oceania, Not Specified, US or territory, Mexico, Philippines, Vietnam, India, Germany, Puerto Rico, El Salvador, Cuba, United Kingdom, Canada, China, South Korea, Taiwan, Guatemala, Japan, Haiti, USSR Core, Jamaica, Colombia, Poland, Iran, Dominican Republic, Italy, Former Socialist Europe, Western Europe. A graduate student is coded as US-born if they have a place of birth value of "US or territory" is coded as foreign-born otherwise. See here for more: https://lehd.ces.census.gov/data/lehd-snapshot-doc/latest/sections/person_level/icf.html#details_icf_us_pob

[6]Though "non-US-born" would be more precise, to ease exposition, we use "foreign-born" with the understanding that "foreign" is relative to the United States.

3.4 Labor Market Outcomes

To track the post-degree labor market outcomes of the UMETRICS Ph.D. recipients, we link them to a variety of confidential data from the Internal Revenue Service (IRS) and the U.S. Census Bureau.[7] In particular, we link them to the universe of W2 tax records (2005-2018) and to the universe of unemployment insurance earnings records (2005-2018) from the Longitudinal Employer-Household Dynamics (LEHD) Program at the U.S. Census Bureau.

W2 Tax Records The Internal Revenue Service (IRS) requires that US employers file form W2, which lists wages paid to and taxes withheld from each employee. The form allows us to track each Ph.D. recipient's annual earnings as well as identify their employer using federal tax identification numbers (EIN). We use the EIN to identify which employers are universities (see IPEDS section below), enabling us to determine whether a Ph.D. recipient is working in a university or non-university job.

Longitudinal Employer-Household Dynamics (LEHD) LEHD data are built using state-level unemployment insurance (UI) earnings records (Vilhuber et al., 2014).[8] These data allow us to track each Ph.D. recipient's quarterly earnings and identify their employer using a state tax identification number (SEIN). Fortunately, most SEINs can be mapped to EINs and so we can use this information to identify university employers and determine whether a Ph.D. recipient's earnings are from a university or non-university (again, see the IPEDS section next).

IPEDS As noted, we link a publicly-available list of EIN tax identifiers for most US-based universities to the W2 and LEHD earnings data, enabling us to determine whether a Ph.D. recipient's post-degree job placements are in academia or industry. The university EIN list is part of the Integrated Postsecondary Education Data System (IPEDS) and is maintained by the National Center for Education Statistics (NCSES).[9]

Attachment to the US Labor Market In addition to using the W2 and LEHD earnings data to track the post-degree labor market outcomes of UMETRICS Ph.D. recipients, we

[7] Using personally identifiable information (PII) such as name and birth date, the Census Bureau assigns individuals to a Protected Identification Key (PIK) which is an internal person-level identifier. This is done through the Person Identification Validation System (PVS), which is a probabilistic match (Wagner et al., 2014). Once an individual is assigned to a PIK, we can link them to a wide variety of confidential data housed at Census.

[8] For the most up-to-date LEHD documentation, see here: https://lehd.ces.census.gov/data/lehd-snapshot-doc/latest/

[9] The EIN data can be found at https://nces.ed.gov/ipeds/datacenter/DataFiles.aspx under the title "Directory Information". We combine the datasets from 2002 to 2018.

also use it to identify subsamples of individuals defined by the strength of their post-degree attachment to the US labor market. We do this because Ph.D. recipients typically have very high employment rates (Mervis, 2016; Milesi et al., 2014) and tend to be quite geographically mobile (CITES). Thus, when our US-based W2/LEHD records indicate that a Ph.D. recipient has zero earnings, there is a high probability they they are in fact working, but outside of the US.[10] Thus, to avoid coding individuals with missing earnings data as zeros, we opt to create conditional-on-positive earnings samples.

Specifically, from the "full sample" of 14,900 UMETRICS Ph.D. recipients, we identify the 12,100 that have positive earnings one year after graduation ("first-year sample") and the 10,550 that have positive earnings in *all four* years after graduation ("4-year sample"). Eliminating observations with (likely) mis-measured zero earnings provides the additional benefit of being able to estimate regression equations using the log of earnings.

Of course, conditioning on positive earnings creates samples of Ph.D. recipients who select into particular levels of attachment to the US labor market and this selection may differ for the US- and foreign-born. For instance, if the foreign-born require a higher wage than their US-born counterparts to stay and work in the United States, then the foreign-born selecting into the conditional-on-positive samples will have higher earnings, biasing upwards our estimates of the foreign-born earnings premium. On the other hand, if the foreign-born are willing to accept lower paying jobs in order to stay in the US, then this may bias our estimates down. Overall, our results show that, as attachment to the US labor market grows stronger, the foreign-born premium rises, (See Tables 1, 3 and 4) suggesting that foreign-born Ph.D. recipients are, in fact, positively selected.

3.5 Analysis Samples

As suggested, we use three main samples for the analyses in this paper. We start with the set of all UMETRICS graduate students that have a Proquest dissertation ("full sample"), which totals 14,900 Ph.D. recipients. Second we construct two subsamples that condition on positive earnings one year post-degree ("first-year sample") and *all four* years post-degree ("4-year sample"), which have 12,100 and 10,550 Ph.D. recipients, respectively.

Summary Statistics Table 1 shows key summary statistics for the Ph.D. recipients in these three samples. Panel A displays labor market outcomes measured one year and four

[10] Tham et al. (2023) use W2/LEHD data combined with the full 2020 Decennial Census to examine the the likelihood that lab personnel leave the US after their lab experiences a delay in the arrival of grant funding. They find that the scientific workforce is, indeed, quite mobile with about half of personnel who receive zero post-delay W2/LEHD earnings also being absent from the 2020 Decennial Census, indicating that they left the US.

years post-degree. In both levels and logs, and across all samples and time-frames, foreign-born Ph.D. recipients earn more than their U.S.-born peers. Remarkably, this is even true in samples and time-frames that allow for labor force detachment, in which case foreign-born Ph.D. recipients are much more likely to be out of the U.S. labor force (and thus have their earnings coded as zero). We also see that, as the strength of attachment to the U.S. labor market increases, so does the foreign-born earnings premium, suggesting that the foreign-born stayers are even more strongly selected in terms of the choices and characteristics that lead to higher earnings. In terms of job placement, the U.S.-born are more likely than their foreign-born peers to be working in academia, less likely to be working in industry, and less likely to be out of the U.S. labor force (when sample construction allows this outcome).

Panel B displays the demographic characteristics of the Ph.D. recipients in our three samples. We see that about a third of Ph.D. recipients in our sample are female and that the U.S. born are about 4-5 pp more likely than the foreign-born to be female. The typical Ph.D. recipient in our sample receives their degree at around age 30 and the U.S.-born are 0.6-0.9 years younger than the foreign-born at graduation. However, there is no difference in when the typical Ph.D. recipient in our samples receives their degree, with 2012 being the average for both the U.S.- and foreign-born.

Finally, Panel C reports on the characteristics of the lab in which a Ph.D. recipient worked while they were in graduate school. The results are quite similar across all samples and time frames. The typical lab has about 1.3-1.4 team members (suggesting that most are quite small) and the U.S.-born tend to work on slightly smaller teams. About half of the personnel in a typical lab are faculty, with the ratio a bit lower for the U.S.-born. The average award that supports a lab is worth about \$350-375k and the U.S.-born tend to be paid by awards that are about \$10-15k larger than their foreign-born peers. Finally, the typical Ph.D. recipient is paid by 2.2-2.5 awards, with the foreign-born being paid my 0.20-0.25 more awards. In sum, the U.S.-born tend to work in labs with slightly larger teams, a lower faculty to other personnel ratio, are paid by fewer awards, and are paid by larger awards.

Field Distribution Figure 1 shows the distribution, over fields, of foreign- and US-born Ph.D. recipients. Most notable, across all three samples, the foreign-born are far more likely than their U.S.-born peers to receive an engineering degree. Indeed, depending on the sample, 48-51% of foreign-born Ph.D. recipients receive an engineering degree compared to 24-27% of the U.S.-born. In contrast, the U.S.-born are more likely to receive a degree in any other field (with the exception of mathematics, where differences are quite small). This is especially true in biology, where 25-29% of U.S.-born Ph.D. recipients receive their degree

in biology compared to about 17% of the foreign-born.

The strong selection of the foreign-born into engineering, which tends to result in highly-paid jobs and the selection of the US-born into biology, which tends to result in relatively low-paying jobs is crucial for understanding the raw earnings premium enjoyed by the foreign-born.[11] Indeed, as we show below using decompositions, field fixed effects account for about 60% of the explained earnings gap between foreign- and US-born Ph.D. recipients.

4 Results

4.1 Lab Regressions

Table 2 displays regressions of various lab characteristics on the foreign-born indicator, which allow us to assess differences in the research environment between the U.S.- and foreign-born while they were in graduate school. Panels A, B, and C display regressions for the full, first-year, and 4-year samples, respectively, which allow us to assess gaps for Ph.D. recipients who are, post-graduation, increasingly strongly attached to the US labor market. The odd columns contain regressions without covariates (except degree-year fixed effects) and the even columns contain regressions with covariates, including a gender indicator, quadratic in age, university fixed effects, and field fixed effects.

For all three samples, column (1) shows that, while in graduate school, the foreign-born were paid by more grants than their U.S.-born counterparts. Indeed, the foreign-born in the full, first-year, and 4-year samples were paid by 6.9%, 8.9% and 9.4% more grants. Column (2) shows that these gaps remain after including covariates, though they are attenuated to 5.0%, 6.8%, and 7.3%. These grant count gaps may reflect that the foreign-born have less stable funding as they bounce from lab to lab, have a more general set of skills that translate across many labs, have a broader research agenda that requires work in multiple labs, or possibly, are more research-oriented compared to the U.S.-born who may be more teaching-focused and thus paid by fewer grants during graduate school. The pattern of increasing coefficients as the sample changes from the full to the first-year to the 4-year suggests that, whatever this difference in the number of grants represents, it is most pronounced among Ph.D. recipients who are more strongly attached to the U.S. labor market.

Column (3) reveals that the foreign-born tend to work in labs supported by smaller grants, with the average grant size (in dollars) being 10.9%, 8.9%, and 6.7% smaller for the foreign-born relative the U.S.-born. Column (4) shows that adding covariates substantially

[11] Using samples of Ph.D. recipients similar to those in this paper, Jiang et al. (2023) find that staring earnings are $98k for engineering graduates and $58K for biology graduates. Four years after graduation, engineer's earnings increased to $124k and biologists to $78k.

attenuates these differences, basically driving the average grant size gap to zero for all three samples. However, the pattern of decreasing (in absolute value) gaps remains as the sample changes from the full to the first-year to the 4-year, suggesting that as the sample of Ph.D. recipients becomes, post-graduation, more strongly attached to the U.S. labor market the difference in grant size between the U.S.- and foreign-born (while they were in graduate school) becomes smaller.

Columns (5) and (6) show that the foreign-born tend to work on smaller teams, but the differences are small and typically statistically insignificant (or marginally significant). These differences range from about 1-2% depending on the sample and are not substantially impacted by the inclusion of covariates. However, whether or not covariates are included, the absolute value of the gaps decrease in size as the sample changes from the full to the first-year to the 4-year, indicating that, among Ph.D. recipients who later become more attached to the U.S. labor market (after graduation), the foreign- and U.S.-born tend to work with more similarly sized teams when they received their graduate training.

Finally, columns (7) and (8) show that the foreign-born tend to work in labs with a higher ratio of faculty to other other personnel, and that, if anything, including covariates increases the difference. Depending on the sample and specification, the foreign-born work in labs with a faculty to other personnel ratio that is about 1.5-2.5% higher than the labs in which the U.S.-born work. Unlike the other three variables measuring the lab environment in which Ph.D. recipients worked while in graduate school, the faculty to other personnel ratio gap does not have a clear pattern as the sample composition changes toward individuals more strongly attached to the US labor market. Indeed, the gaps are largest for Ph.D. recipients in the first-year sample with smaller (and quite similar) gaps for the full and 4-year sample.

In sum, while they were in graduate school, foreign-born Ph.D. recipients tended to be paid by more grants, worked on smaller grants, worked on smaller teams, and worked on teams with more faculty relative to other personnel. Though it is not clear whether higher/lower values on these outcomes are necessarily "better", it does seem clear that the foreign-born select into *different* types of research environments than their U.S.-born counterparts when they are in graduate school. In particular, it appears that the foreign-born select into a more intimate lab environment, where the grants and teams are smaller and there are more faculty per teammate.

As noted, there are also interesting patterns that emerge as the sample composition changes toward Ph.D. recipients who exhibit stronger attachment to the U.S. labor market. In terms of award count, the foreign- and U.S.-born have a less similar research environment as the sample moves from the full to the first-year to the 4-year. In terms of average award size and team size, the foreign- and U.S.-born have a more similar research environment as

the sample composition changes toward Ph.D. recipients who are more strongly attached to the U.S. labor market after graduation. There is not a monotonic pattern for the faculty ratio.

4.2 Earnings Regressions

In the previous section, we showed that, while they were in graduate school, foreign-born Ph.D. recipients select into different types of lab environments than their U.S.-born peers. In this section, we examine the selectivity of the foreign-born using post-gradation career outcomes.

Table 3 displays regressions of log earnings outcomes on the foreign-born indicator, which reveals the foreign-born earnings premium. Columns (1)-(2) display regressions for the first-year sample and columns (3)-(6) display regressions for the 4-year sample. The odd columns contain regressions without covariates (except degree-year fixed effects) and the even columns contain regressions with covariates, including a gender indicator, quadratic in age, measures of lab characteristics while the Ph.D. recipient was in graduate school, university fixed effects, field fixed effects, and an indicator for whether the Ph.D. recipient is currently working in an academic position (as opposed to working in an industry position).

Column (1) shows that, for Ph.D. recipients with positive earnings in their first year after graduate school (first-year sample), the raw earnings gap between the foreign- and US-born is 14.4%. This is direct evidence that the foreign-born are strongly positively selected in the sense that they possess characteristics (e.g. ability) and make choices (e.g. field of study or academic/industry job) that lead to substantially higher earnings than their US-born peers. Even after adding covariates the foreign-born earnings premium remains, though it is substantially attenuated at 2.3%. A Gelbach decomposition reveals that 61.6% of the explained log earnings gap is accounted for by field fixed effects and 37.6% is accounted for by the academia indicator.[12] In total, these two covariates account for practically all (99%) of the explained portion of the foreign-born earnings premium. This conditional gap suggests that the foreign-born possess characteristics (again, perhaps ability) that lead to higher earnings even when choices like field of study and job type are accounted for. Interestingly, the Gelbach decomposition suggests that the university fixed effects expand the log earnings gap (though, by a modest 3%), suggesting that the foreign-born tend to graduate from lower ranked universities (or at least universities that tend to place graduates in lower paying jobs).

Columns (3)-(6) tell a similar story for Ph.D. recipients who have positive earnings, not

[12]By "explained" log earnings gap, we mean the difference between the raw and conditional regression coefficients on the foreign-born indicator: 0.1435 - 0.02279 = 0.12071, or 12.1%. The "unexplained" gap is simply the remaining gap after observable covariates have been accounted for: 0.02279, or 2.3%.

only in their first year after graduate school, but also have positive earnings for *all* years 1-4 after receiving their degree. At one and four years post-degree, the raw foreign-born earnings premia are 17.0% and 17.6%. The premia at both time horizons remain after controlling for covariates, but are again substantially attenuated to 4.5% and 6.1%. As before, field fixed effects and the academia indicator play the primary roles in attenuating these log earnings gaps, together accounting for 99% at one year post-degree and 106% at four years post-degree.[13]

It is notable that the foreign-born earnings premium in the first year post-degree is larger for the 4-year sample than for the first-year sample. This suggests that the Ph.D. recipients who are more strongly attached to the U.S. labor market, and end up working in the U.S. for all four years post-degree, are more positively selected than those who stay for only one year. That is, the foreign-born who exit the U.S. labor force, likely returning to their home country, tend to earn less in the U.S. than the foreign-born graduates who are more strongly attached to the U.S. labor market. For the 4-year sample, it is notable that both the raw and conditional log earnings gaps are larger in year 4 than in year 1. This suggests that, not only do the foreign-born earn more immediately after graduating, but their earnings grow at a faster rate than their US-born counterparts.

If we consider the traditional omitted variable bias (OVB) formula, unobservable covariates such as "ability" are quite likely to be positively related to earnings (even conditional on observed covariates). In the likely case that omitting ability from our regressions biases our estimates of the foreign-born earnings premium upward, this implies that ability must be positively related to the foreign-born indicator. That, it is very likely that the foreign-born are positively selected, not only on observed covariates like field, but also unobservable covariates like "ability".

4.3 Sensitivity: Selection on Unobservable Covariates

As noted, reasoning from the OVB formula suggests that it is quite likely that the foreign-born are positively selected in terms of unobservable covariates like ability. In this section, we attempt to quantify the strength of this selection using two metrics proposed by Cinelli and Hazlett (2020) – the "robustness value" (RV) and the "extreme scenario value" (ESV) – that use the OVB formula to assess the sensitivity of linear regression coefficients to unobservable covariates.

These sensitivity statitsics are displayed in Table 3. The RV measures how strong the

[13]Variables can account for over 100% if other variables expand the log earnings gap between US- and foreign-born Ph.D. recipients. For instance, university fixed effects again expand the log earnings gaps, suggesting that the foreign-born attend lower ranked schools.

residual associations between an unobservable covariate (such as ability) and *both* log earnings *and* the foreign-born indicator would have to be in order to drive the foreign-born earnings premium to zero.[14] The ESV measures how strong the residual association between an unobservable covariate and the foreign-born indicator would have to be to eliminate the foreign-born earnings premium in the extreme case where the unobservable covariate explains *all* of the residual variation in the outcome.[15]

The RV of the foreign-born coefficient for year-1 earnings in the first-year sample is 0.0212, meaning that if an unobservable covariate such as ability could explain at least 2.1% of the residual variance of *both* log earnings and the foreign-born indicator, it would drive the foreign-born earnings premium to zero. The ESV is a much smaller 0.00046, meaning that in the extreme scenario where an unobservable covariate explained *all* of the residual variation in year-1 log earnings, then it would only need to explain 0.046% of the residual variation in the foreign-born indicator to eliminate the the foreign-born earnings premium. For the 4-year sample, the RVs of the foreign-born coefficient are a bit larger at 4.4% and 5.6% for year-1 and year-4 log earnings, respectively. The ESVs are also larger at 0.20% and 0.33%.

As discussed, we find it likely that, in the absence of selection, the foreign-born earnings premium would indeed vanish. If we believe that "ability" is the primary unobservable covariate and that it likely explains most (or all) of the residual variation in log earnings, then we can view the ESV values as loose upper bounds on the "ability gap" between the US- and foreign-born. The ESV values of 0.046%, 0.20%, and 0.33% suggests that these conditional ability gaps can be quite small and still eliminate the foreign-born earnings premium. Indeed, as discussed below, all observed covariates yield values strong enough that, if ability had the same strength, it would eliminate the earnings gap. Thus, selection on unobservable ability appears not to be that strong and most selection is captured by observable covariates such as field and sector of employment. It is also notable that the ESVs are substantially larger (4-7 times) for the 4-year sample than for the 1-year sample.

[14] By "residual association" or "residual variation", we mean partial R^2 values. Using the notation of Cinelli and Hazlett (2020), let Y be the outcome in a linear regression (log earnings in our case), D be the "treatment" variable (foreign-born indicator in our case), X be a set of observable covariates, and Z be an unobservable covariate. Then the residual association, or partial R^2, of the unobservable covariate Z with the outcome Y (conditional on the treatment D and observable covariates X) is denoted $R^2_{Y\sim Z|D,X}$. This quantity would be obtained by regressing Y on D and X and computing the OLS residuals, regressing Z (if we could observe it) on D and X and computing the OLS residuals, and finally obtaining the R^2 value from regressing the first set of residuals on the second set of residuals. Similarly, the residual association of the unobservable covariate with the treatment (conditional on observable covariates) is denoted $R^2_{D\sim Z|X}$. This quantity would be obtained by regressing D on X and regressing Z on X, obtaining the OLS residuals for both regressions, and obtaining the R^2 value from a regression of the first set of residuals on the second set of residuals.

[15] That is, in the extreme scenario where $R^2_{Y\sim Z|D,X} = 1$.

This is further evidence that the foreign-born with strong attachment to the US labor market are even more selected than those with less strong attachment.

Is it plausible that there exists an unobservable covariate like ability that could explain this much residual variation in log earnings and the foreign-born indicator? To assess this, we use methods from Cinelli and Hazlett (2020) to benchmark the RVs and ESVs for each sample and log earnings outcome against the residual variation that *observed* covariates *would* explain *if* they had been omitted from the regression. These benchmark residual variation values are displayed in Table 4, which also displays regression output for different specifications where each set of observed covariates are, one at a time, omitted from the regression. Figure 2 organizes the RVs (dashed vertical lines) and ESVs (solid vertical lines) as well as benchmark values of the observed covariates' associations with log earnings (red bars) and the foreign-born indicator (blue bars). When an observed covariate has partial R^2 values that are *both* ($R^2_{Y\sim X_j|D,X_{-j}}$ and $R^2_{D\sim X_j|X_{-j}}$) above the corresponding RV, it implies that an unobservable covariate with the same strength would bring the earnings gap to zero. When an observed covariate has a partial R^2 with the foreign-born indicator ($R^2_{D\sim X_j|X_{-j}}$) that is above the corresponding ESV, it implies that an unobservable covariate with the same strength would eliminate the foreign-born earnings premium in the extreme scenario in which the unobservable covariate explained all of the residual variation of the log earnings outcome.

For the first-year sample, we see that field fixed effects explain 5.7% of the residual variation in the foreign-born indicator and 8.7% of the residual variation in year-1 log earnings. Since 5.7% is above the corresponding ESV of 0.046%, an unobserved covariate as strongly related to the foreign-born indicator as field fixed effects would eliminate the foreign-born earnings premium in the extreme scenario where the unobservable covariate explained all of the residual variation in year-1 log earnings. Moreover, beyond this extreme scenario, since 5.7% and 8.7% are both above the corresponding RV of 2.1%, an unobservable covariate as strongly related to *both* year-1 log earnings *and* the foreign-born indicator as field fixed effects would also drive the foreign-born earnings premium to zero.

In contrast, we see that lab characteristics explain only 0.53% and 0.13% of the residual variations in the foreign-born indicator and year-1 log earnings. Since these are both below the RV of 5.7%, an unobservable covariate as strongly related to the foreign-born indicator and year-1 log earnings would *not* eliminate the foreign-born earnings premium. However, since 0.53% is above the corresponding ESV of 0.046%, an unobserved covariate as strongly related to the foreign-born indicator as lab characteristics would eliminate the foreign-born earnings premium *if* the unobservable covariate explained all of the residual variation in year-1 log earnings.

Indeed, in this extreme scenario an unobservable covariate as strong as *any* of our observed covariates would be strong enough to eliminate the foreign-born earnings gap. However, absent this extreme scenario, it is ambiguous whether unobservable covariates as strong as the remaining observed covariates – demographic characteristics, sector of employment, and university fixed effects – would eliminate the premium. For these, one residual association is above the RV and the other is below. Indeed, sector is very strongly related to year-1 log earnings but weakly related to the foreign-born indicator whereas demographic characteristics and university fixed effects are strongly related to the foreign-born indicator and only weakly related to year-1 log earnings.

Table 4 and Figure 2 show that the results are very similar for both year-1 and year-4 log earnings for Ph.D. recipients in the 4-year sample. All observed covariates have high enough residual associations with the foreign-born indicator that, if an unobservable covariate had the same strength, it would eliminate the foreign-born earnings premium in the extreme scenario where the unobservable explained all of the residual variation in log earnings. Beyond this extreme scenario, we again see that only field fixed effects have large enough residual associations with both log earnings and the foreign-born indicator that, if an unobservable covariate had the same associations, it would bring the foreign-born earnings premium to zero. It also remains ambiguous whether unobservable covariates as strong sector would eliminate the premium (again, sector is strongly related to log earnings but weakly related to the foreign-born indicator). Unlike the first-year sample, unobservable covariates as strong as the demographic covariates and university fixed effects would not eliminate the foreign-born earnings premium because the residual associations with the foreign-born indicator are not above the corresponding RVs.

Overall, these sensitivity measures suggest that, conditional on observed covariates, an unobservable covariate such as ability would not need to be as strong as field fixed effects but would need to be stronger than all other observable covariates to eliminate the foreign-born earnings premium. If we believe ability explains *all* of the residual variation in log earnings, then all covariates with the exception of lab characteristics in the 4-year sample would be strong enough to eliminate the premium.

Ability is almost surely related to log earnings, but there is little reason to believe it is strongly related to the birthplace of a US-trained Ph.D. recipient. However, for all samples and outcomes, the ESV suggests that that, in the extreme case where the omitted variables explain all of the residual variation in the outcome, they need only explain a very small part of the residual variation in the foreign-born indicator to eliminate the result. All variables except training in the 4-year sample exceed this threshold.

5 Conclusion

In this paper, we have examined the extent to which foreign-born Ph.D. recipients are positively selected, relative to their foreign-born peers, in terms of the observed choices they make (e.g. type of lab, field, or sector of employment) as well as their unobserved characteristics (e.g. "ability"). Our findings reveal a positive selection process among foreign-born Ph.D. recipients, suggesting that the U.S. retains the "best and brightest" in economically critical fields. Importantly, our research demonstrates that foreign-born Ph.D. recipients tend to study in fields with high demand and better economic prospects, such as engineering. This not only contributes to the overall positive selection but also suggests minimal crowding out of domestic trainees in these lucrative fields.

The examination of post-graduation outcomes reveals a small but persistent earnings premium for foreign-born Ph.D. recipients, even after controlling for field of study and various other factors. Furthermore, our results indicate a growing earnings premium over time for foreign-born doctoral recipients. This positive trend suggests that foreign-born STEM Ph.D.s are not only positively selected, but they also experience faster earnings growth than their US-born counterparts.

References

Amuedo-Dorantes, Catalina and Delia Furtado (2019), "Settling for academia? h-1b visas and the career choices of international students in the united states." *Journal of Human Resources*, 54, 401–429. 5

Brown, J. David, Misty L. Heggeness, Suzanne M. Dorinski, Lawrence Warren, and Moises Yi (2018), "Understanding the Quality of Alternative Citizenship Data Sources for the 2020 Census." Working Papers 18-38, Center for Economic Studies, U.S. Census Bureau, URL https://ideas.repec.org/p/cen/wpaper/18-38.html. 7

Cinelli, Carlos and Chad Hazlett (2020), "Making sense of sensitivity: Extending omitted variable bias." *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 82, 39–67. 14, 15, 16, 26

Ganguli, Ina and Patrick Gaulé (2019), "Will the US Keep the Best and the Brightest (as Postdocs)? Career and Location Preferences of Foreign STEM PhDs." In *The Roles of Immigrants and Foreign Students in US Science, Innovation, and Entrepreneurship*, NBER Chapters, 49–69, National Bureau of Economic Research, Inc, URL https://ideas.repec.org/h/nbr/nberch/14108.html. 4

Grogger, Jeffrey and Gordon H. Hanson (2015), "Attracting Talent: Location choices of foreign-born PhDs in the United States." *Journal of Labor Economics*, 33, S5–S38. 4, 5

Hunt, Jennifer and Marjolaine Gauthier-Loiselle (2010), "How much does immigration boost innovation?" *American Economic Journal: Macroeconomics*, 2, 31–56, URL https://www.aeaweb.org/articles?id=10.1257/mac.2.2.31. 2

Ikudo, Akina, Julia I Lane, Joseph Staudt, and Bruce A Weinberg (2019), "Occupational classifications: A machine learning approach." *Journal of Economic and Social Measurement*, 44, 57–87. 6

Jiang, Xuan, Joseph Staudt, and Bruce A Weinberg (2023), "A tale of two fields? stem career outcomes." Working Paper 31835, National Bureau of Economic Research, URL http://www.nber.org/papers/w31835. 11

Kahn, Shulamit and Megan MacGarvie (2020), "The impact of permanent residency delays for STEM PhDs: Who leaves and why." *Research Policy*, 49, 103879, URL http://www.nber.org/papers/w25175https://linkinghub.elsevier.com/retrieve/pii/S0048733319301982. 5

Kerr, William R. and William F. Lincoln (2010), "The Supply Side of Innovation: H-1B Visa Reforms and U.S. Ethnic Invention." *Journal of Labor Economics*, 28, 473–508, URL https://www.journals.uchicago.edu/doi/10.1086/651934. 2

Khosla, Pooja (2018), "Wait time for permanent residency and the retention of immigrant doctoral recipients in the U.S." *Economic Analysis and Policy*, 57, 33–43. 5

Lan, Xiaohuan (2012), "Permanent Visas and Temporary Jobs: Evidence from Post-doctoral Participation." *Journal of Policy Analysis and Management*, 31, 623–640, URL http://web.a.ebscohost.com/ehost/pdfviewer/pdfviewer?vid=2{&}sid=899ea6c6-dc60-495b-88f7-0cb7b190dcfd{%}40sessionmgr4008. 5

Lane, Julia I, Jason Owen-Smith, Rebecca F Rosen, and Bruce A Weinberg (2015), "New linked data on research investments: Scientific workforce, productivity, and public value." *Research Policy*, 44, 1659–1671. 6

Mervis, Jeffrey (2016), ""employment crisis" for new ph.d.s is an illusion." *Science*, URL http://dx.doi.org/10.1126/science.caredit.a1600082. 9

Milesi, Carolina, Lance A Selfa, and Lynn M Milan (2014), "Unemployment among doctoral scientists and engineers increased but remained below the national average." *Psychology*, 89, 93–000. 9

Nicholls, Natsuko, Christopher Brown, and Jason Owen-Smith (2022), "Data dictionary for IRIS UMETRICS 2022 data release." 6

ProQuest (2018). Dissertations Theses Database, URL https://about.proquest.com/en/dissertations/. 6

Roach, Michael, Henry Sauermann, and John Skrentny (2019), "Are foreign stem phds more entrepreneurial? entrepreneurial characteristics, preferences, and employment outcomes of native and foreign science and engineering phd students." In *The Roles of Immigrants and Foreign Students in US Science, Innovation, and Entrepreneurship*, 207–228, University of Chicago Press. 5

Tham, Wei Yang, Joseph Staudt, Elisabeth Ruth Perlman, and Stephanie Cheng (2023), "The effect of funding delays on the research workforce: Evidence from tax records." Technical report, Mimeo. 9

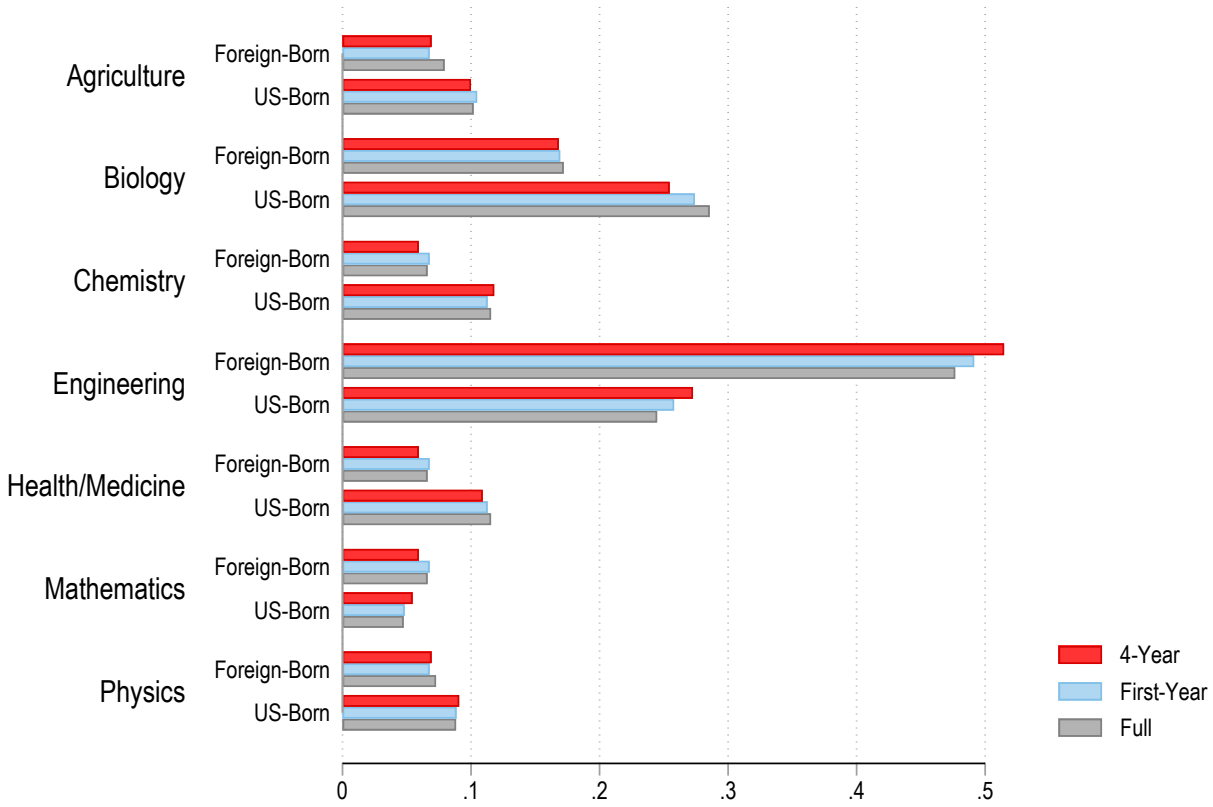
UMETRICS (2018, Q4). Institute for Research on Innovation and Science (IRIS), University of Michigan, URL https://iris.isr.umich.edu/. 6

Vilhuber, Lars, Kevin McKinney, et al. (2014), "Lehd infrastructure files in the census rdc-overview." *Center for Economic Studies, US Census Bureau Working Papers*. 7, 8

Wagner, Deborah, Mary Lane, et al. (2014), "The person identification validation system (pvs): applying the center for administrative records research and applications'(carra) record linkage software." Technical report, Center for Economic Studies, US Census Bureau. 8

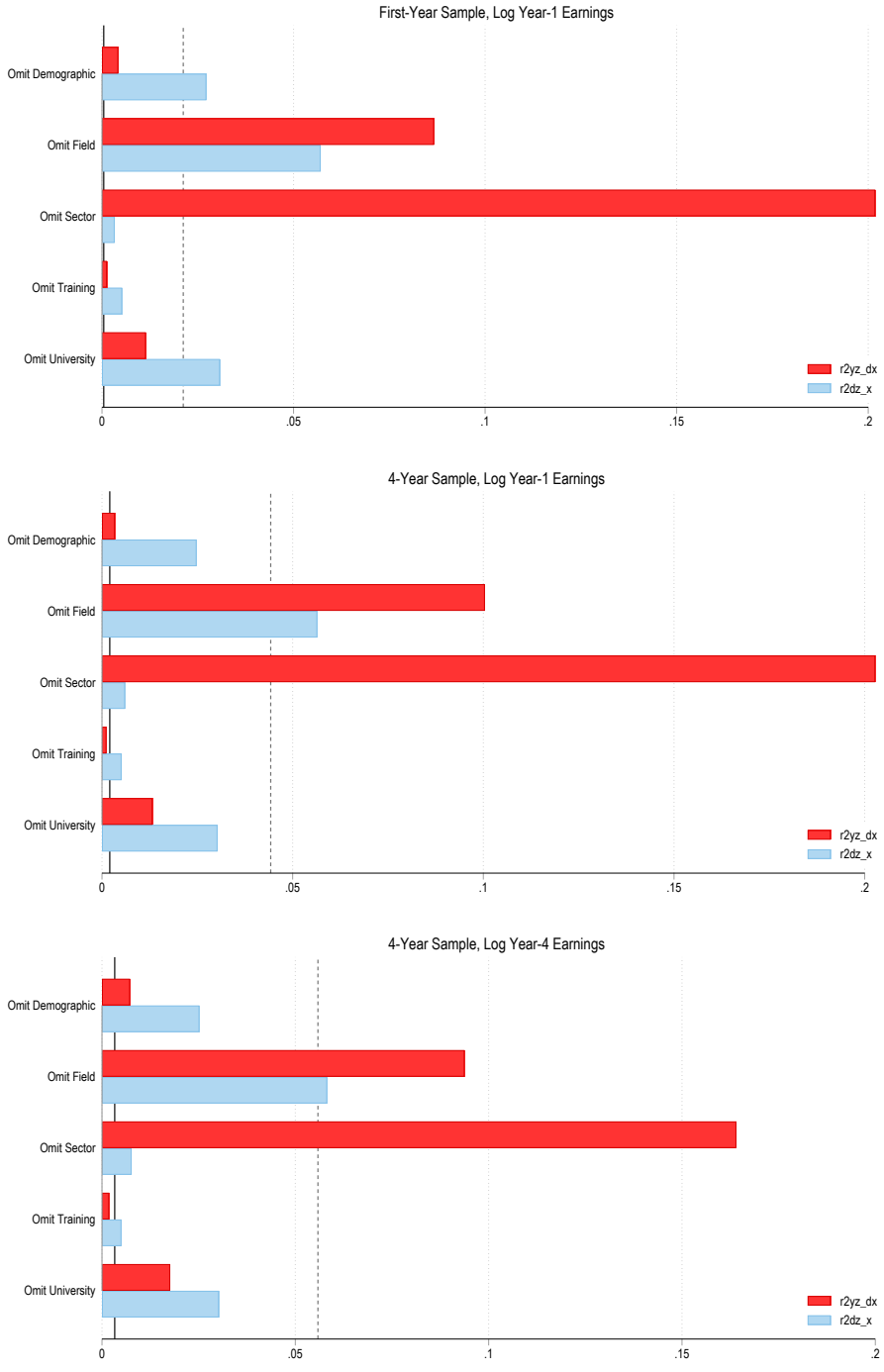
Zeithammer, Robert and Ryan P. Kellogg (2013), "The hesitant hai gui: Return-migration preferences of u.s.-educated chinese scientists and engineers." *Journal of Marketing Research*, 50, 644–663, URL https://doi.org/10.1509/jmr.11.0394. 4

Figure 1: Field Distribution of Foreign-Born and US-Born Ph.D. Recipients



Notes – This figure shows the distributions, over field, of foreign-born and US-born Ph.D. recipients for our three main samples. Thus, for each place of birth category and each sample, the fractions summed over the seven fields equal 1. The "Full" sample includes all graduate students in UMETRICS that can be linked to their dissertation in ProQuest. The "First-Year" and "4-Year" samples include the subset of graduate students from the Full sample that have, respectively, positive earnings the first year after receiving their Ph.D. and positive earnings in all years 1-4 after receiving their Ph.D.

Figure 2: Sensitivity Benchmarks Based on Observable Covariates



Notes – .

Table 1: Summary Statistics

	Full Sample			First Year Sample			All Four Years Sample		
	US	**Foreign**	**Diff.**	**US**	**Foreign**	**Diff.**	**US**	**Foreign**	**Diff.**
Panel A: Labor Market Outcomes									
Earnings (Year 1)	58,320	65,110	-6,790***	69,490	83,100	-13,610***	72,370	88,410	-16,040***
	(44,860)	(58,370)		(40,270)	(53,420)		(40,560)	(54,400)	
Earnings (Year 4)	85,760	89,590	-3,830***	92,350	109,600	-17,250***	100,100	127,000	-26,900***
	(69,440)	(101,800)		(70,990)	(104,600)		(69,790)	(103,100)	
Log Earnings (Year 1)				11.00	11.14	-0.14***	11.05	11.22	-0.17***
				(0.56)	(0.62)		(0.53)	(0.59)	
Log Earnings (Year 4)							11.37	11.54	-0.17***
							(0.53)	(0.64)	
Academic Placement (Year 1)	0.39	0.29	0.10***	0.47	0.37	0.10***	0.45	0.33	0.12***
	(0.49)	(0.45)		(0.50)	(0.48)		(0.50)	(0.47)	
Academic Placement (Year 4)	0.36	0.21	0.15***	0.37	0.24	0.13***	0.38	0.27	0.11***
	(0.48)	(0.41)		(0.48)	(0.43)		(0.49)	(0.45)	
Industry Placement (Year 1)	0.45	0.49	-0.05***	0.53	0.63	-0.10***	0.55	0.67	-0.12***
	(0.50)	(0.50)		(0.50)	(0.48)		(0.50)	(0.47)	
Industry Placement (Year 4)	0.55	0.52	0.02***	0.58	0.63	-0.06***	0.62	0.73	-0.11***
	(0.50)	(0.50)		(0.49)	(0.48)		(0.49)	(0.45)	
Out of Labor Force (Year 1)	0.16	0.22	-0.06***	0.00	0.00	0.00	0.00	0.00	0.00
	(0.37)	(0.41)		(0.00)	(0.00)		(0.00)	(0.00)	
Out of Labor Force (Year 4)	0.09	0.26	-0.17***	0.06	0.13	-0.07***	0.00	0.00	0.00
	(0.29)	(0.44)		(0.23)	(0.33)		(0.00)	(0.00)	
Panel B: Demographic Characteristics									
Female	0.36	0.31	0.05***	0.35	0.30	0.05***	0.34	0.30	0.04***
	(0.48)	(0.46)		(0.48)	(0.46)		(0.47)	(0.46)	
Age	29.38	30.28	-0.90***	29.40	30.09	-0.69***	29.41	30.03	-0.62***
	(2.81)	(3.04)		(2.84)	(2.96)		(2.85)	(2.96)	
Degree Year	2012	2012	0.00	2012	2012	0.00	2012	2012	0.00
	(3.00)	(3.06)		(3.00)	(3.04)		(2.99)	(3.03)	
Panel C: Lab Characteristics									
Team Size	1.38	1.37	0.01	1.36	1.33	0.04	1.38	1.36	0.02
	(2.26)	(3.73)		(2.20)	(2.16)		(2.23)	(2.20)	
Faculty Ratio	0.48	0.50	-0.02*	0.48	0.50	-0.03**	0.48	0.50	-0.02
	(0.68)	(0.77)		(0.68)	(0.80)		(0.70)	(0.82)	
Awards	2.24	2.44	-0.20***	2.25	2.50	-0.25***	2.26	2.52	-0.26***
	(1.52)	(1.68)		(1.54)	(1.72)		(1.56)	(1.74)	
Award Amount	366,800	351,000	15,800	365,000	353,800	11,200	372,900	359,900	13,000
	(945,000)	(940,700)		(967,600)	(1,0040,000)		(1,008,000)	(987,900)	
Arcsine Team Size	0.86	0.84	0.02*	0.86	0.84	0.01	0.86	0.85	0.01
	(0.66)	(0.65)		(0.66)	(0.64)		(0.66)	(0.65)	
Arcsine Faculty Ratio	0.40	0.42	-0.02**	0.40	0.42	-0.02***	0.40	0.42	-0.02**
	(0.42)	(0.40)		(0.42)	(0.40)		(0.42)	(0.40)	
Arcsine Awards	1.40	1.47	-0.07***	1.41	1.49	-0.09***	1.41	1.50	-0.09***
	(0.56)	(0.57)		(0.56)	(0.57)		(0.57)	(0.57)	
Arcsine Award Amount	12.67	12.57	0.10***	12.66	12.58	0.08***	12.67	12.60	0.07**
	(1.47)	(1.67)		(1.47)	(1.65)		(1.47)	(1.63)	
Ph.D. Recipient Count	7,350	7,550		6,200	5,900		5,500	5,050	

Notes – This table contains means and standard deviations (in parentheses below the corresponding mean), for each of our three main samples, separately for foreign-born and US-born Ph.D. recipients. It also contains the differences of the means between the US- and foreign-born (*, **, and *** denote statistical significance at the 10%, 5%, and 1% levels). The variables include labor market outcomes such as earnings, log earnings, and sector of job placement one and four years after a Ph.D. recipient receives their degree. The variables also include demographic characteristics such as sex and age. Finally, the variables include degree year and various characteristics of the labs in which Ph.D. recipients worked while they were in graduate school such as team size, the faculty ratio, the number of awards, and the award amount. The "Full" sample includes all graduate students in UMETRICS that can be linked to their dissertation in ProQuest. The "First-Year" and "4-Year" samples include the subset of graduate students from the Full sample that have, respectively, positive earnings the first year after receiving their Ph.D. and positive earnings in all years 1-4 after receiving their Ph.D.

Table 2: Lab Regressions

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Full Sample								
	Arcsine Award Count		Arcsine Award Amount		Arcsine Team Size		Arcsine Faculty Ratio	
Foreign-Born Indicator	0.06938***	0.05017***	-0.1092***	-0.02131	-0.02196**	-0.02177*	0.01357**	0.01968***
	(0.009276)	(0.009442)	(0.0259)	(0.0199)	(0.01074)	(0.01151)	(0.006728)	(0.006951)
Ph.D. Recipient Count	14,900	14,900	14,900	14,900	14,900	14,900	14,900	14,900
Panel B: First Year Sample								
	Arcsine Award Count		Arcsine Award Amount		Arcsine Team Size		Arcsine Faculty Ratio	
Foreign-Born Indicator	0.08937***	0.06756***	-0.08861***	-0.01843	-0.01858	-0.02089*	0.01913**	0.02548***
	(0.01035)	(0.01049)	(0.02846)	(0.02184)	(0.01182)	(0.01247)	(0.007451)	(0.007657)
Ph.D. Recipient Count	12,100	12,100	12,100	12,100	12,100	12,100	12,100	12,100
Panel C: All Four Years Sample								
	Arcsine Award Count		Arcsine Award Amount		Arcsine Team Size		Arcsine Faculty Ratio	
Foreign-Born Indicator	0.09396***	0.07249***	-0.06708**	-0.003738	-0.01279	-0.01333	0.01472*	0.01961**
	(0.0111)	(0.01121)	(0.03028)	(0.02346)	(0.01277)	(0.01343)	(0.007981)	(0.008192)
Ph.D. Recipient Count	10,550	10,550	10,550	10,550	10,550	10,550	10,550	10,550
Year FEs	×	×	×	×	×	×	×	×
Covariates		×		×		×		×

Notes – This table shows ...

Table 3: Earnings Regressions

	(1)	(2)	(3)	(4)	(5)	(6)
	First Year Sample		**All Four Years Sample**			
	Log Yr-1 Earnings		Log Yr-1 Earnings		Log Yr-4 Earnings	
Foreign-Born Indicator	0.1435***	0.02279**	0.1701***	0.04529***	0.1759***	0.06149***
	(0.0108)	(0.009614)	(0.01096)	(0.0097)	(0.01149)	(0.0105)
Gelbach Percent Contributions						
Field FEs		61.63		58.11		63.75
Academic Placement		37.55		40.91		42.57
Demographic Covariates		1.20		1.09		-2.61
Training		2.70		2.33		2.68
University FEs		-3.09		-2.45		-6.39
Ph.D. Recipient Count	12,100	12,100	10,550	10,550	10,550	10,550
R-Squared	0.0228	0.3043	0.0312	0.3187	0.0328	0.2895
RV		0.0212		0.0442		0.0559
ESV		0.00046		0.00204		0.00329
Year FEs	×	×	×	×	×	×
Covariates		×		×		×

Notes – This table contains coefficients and standard errors (in parentheses below the corresponding coefficient) from regressions of log earnings on an indicator for whether the Ph.D. recipient is foreign-born (*, **, and *** denote statistical significance at the 10%, 5%, and 1% levels). The odd columns are the raw earnings gaps (conditional on degree year fixed effects) and the even columns include covarariates. "Log Yr-1 Earnings" and "Log Yr-4 Earnings" are the log earnings of Ph.D. recipients one year and four years after they receive their degree. The table also includes the percent contributions from Gelbach demopositions, showing which groups of variables play the most important role in attenuating the raw regression coefficients in the odd columns to the conditional coefficients in the even columns. "Field FEs" include indicators for each of the seven fields, "Academic Placement" is an indicator for whether the Ph.D. recipient is working at a university, "Demographic Covariates" include an indicator for gender and a quadratic in age, "Training" includes the arcines of the number of awards, dollar amount of awards, and the faculty ratio for the labs in which the Ph.D. recipient worked while they were in gradaute school, and "University FEs" includes indicators for each of the XX universities. The "First-Year" and "4-Year" samples include all graduate students in UMETRICS that can be linked to their dissertation that have, respectively, positive earnings the first year after receiving their Ph.D. and positive earnings in all years 1-4 after receiving their Ph.D. The rows labeled "RV" and "ESV" are the "robustness value" and "extreme scenario value", which measure the sensitivity of the estimates to unobservable covariates (Cinelli and Hazlett, 2020). For instance, the ESV of 0.00046 means that in the case where unobservable covariates explain *all* of the residual variation in earnings, then these covariates would need to explain at least 0.046 percent of the residual variance of foreign-born status to drive the earnings gap to zero.

Table 4: Sensitivity: Omitting Covariates from Earnings Regressions

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A: First Year Sample							
Dep. Var: Log Year-1 Earnings							
Foreign-Born Indicator	0.02279**	0.01778*	0.02531***	0.05025***	0.08546***	0.01966**	0.1435***
	(0.009614)	(0.009494)	(0.0096)	(0.01053)	(0.009751)	(0.009561)	(0.0108)
Outcome R-Squared	0.3043	0.2968	0.3034	0.1648	0.2505	0.3015	0.02282
Treatment R-Squared	0.1267	0.09981	0.1221	0.1238	0.07685	0.1029	0.002664
Oster's δ	0.8005	0.4875	-1.9252	-0.1142	-0.7876	-0.2756	
Unobservable (Z) as strong as omitted observable (bounds):							
$R^2_{Y\sim Z\|D,X}$	0.0000	0.0115	0.0013	0.2019	0.0867	0.0043	0.5386
$R^2_{D\sim Z\|X}$	0.0000	0.0308	0.0053	0.0033	0.0571	0.0273	0.1420
Ph.D. Recipient Count	12,100	12,100	12,100	12,100	12,100	12,100	12,100
Panel B: All Four Years Sample							
Dep. Var: Log Year-1 Earnings							
Foreign-Born Indicator	0.04529***	0.04052***	0.04762***	0.08016***	0.1058***	0.04304***	0.1701***
	(0.0097)	(0.009609)	(0.009688)	(0.01065)	(0.009879)	(0.009653)	(0.01096)
Outcome R-Squared	0.3187	0.3102	0.3179	0.1822	0.2576	0.3165	0.03122
Treatment R-Squared	0.1257	0.09922	0.1212	0.1204	0.07633	0.104	0.002572
Unobservable (Z) as strong as omitted observable (bounds):							
$R^2_{Y\sim Z\|D,X}$	0.0000	0.0133	0.0012	0.2028	0.1004	0.0034	0.5603
$R^2_{D\sim Z\|X}$	0.0000	0.0303	0.0051	0.0061	0.0565	0.0248	0.1408
Ph.D. Recipient Count	10,550	10,550	10,550	10,550	10,550	10,550	10,550
Panel C: All Four Years Sample							
Dep. Var: Log Year-4 Earnings							
Foreign-Born Indicator	0.06149***	0.05264***	0.06451***	0.0989***	0.1238***	0.05115***	0.1759***
	(0.0105)	(0.01042)	(0.01048)	(0.01127)	(0.01066)	(0.01037)	(0.01149)
Outcome R-Squared	0.2895	0.2777	0.2882	0.1747	0.2302	0.2846	0.03282
Treatment R-Squared	0.1271	0.1007	0.1227	0.1204	0.07626	0.1051	0.002572
Unobservable (Z) as strong as omitted observable (bounds):							
$R^2_{Y\sim Z\|D,X}$	0.0000	0.0176	0.0018	0.1641	0.0938	0.0073	0.4815
$R^2_{D\sim Z\|X}$	0.0000	0.0302	0.0050	0.0077	0.0582	0.0252	0.1427
Ph.D. Recipient Count	10,550	10,550	10,550	10,550	10,550	10,550	10,550
Include All Covariates (Full Model)	×						
Omit University FEs		×					
Omit Training Covariates			×				
Omit Academic Placement Indicator				×			
Omit Field FEs					×		
Omit Demographic Covariates						×	
Omit All Covariates (Except Year FEs)							×

Notes – This table shows ...