

“Estimating Stay Rates for Foreign Recipients of U.S. Science and Engineering Degrees”

A paper prepared for the workshop on
Career Patterns of Foreign-Born Scientists
November 7, 2007

at

National Bureau of Economic Research
Cambridge, MA

prepared by

Michael Finn
Oak Ridge Institute for Science and Education
P.O Box 117
Oak Ridge, TN 37831-0117

This document was prepared by the Oak Ridge Institute for Science and Education (ORISE) through an interagency agreement with the U.S. Department of Energy (DOE). ORISE is managed by Oak Ridge Associated Universities under DOE contract number DE-AC05-06OR23100.

Introduction

In this paper I summarize my research on estimating the stay rate of foreign citizens receiving science or engineering (S/E) doctorates from U.S. universities. Similar estimates of the stay rate for foreign citizens receiving S/E bachelors and masters do not exist. However, I briefly discuss two possible approaches for using tax return data to producing such estimates.

Stay Rates of Doctorate Recipients

My preferred method for estimating stay rates has been to use the Doctorate Record File (DRF), the repository of years of SED data, only to identify foreign doctorate recipients. Then, the Social Security Master Earnings File is used to estimate how many of the foreign doctorate recipients paid taxes on U.S. earnings in any given year. This works because the SED collects Social Security numbers (SSNs) from respondents, unlike most of the other NSF-sponsored surveys. The challenge was to do this in a way that minimizes errors and does not violate the confidentiality of the individuals involved.

To ensure confidentiality of data supplied by doctorate recipients, I have no access to confidential data. Instead, the National Opinion Research Center (NORC) sends groups of Social Security numbers (SSNs) to the U.S. Social Security Administration (SSA). For example, one group was temporary residents earning economics degrees during 2001. Then SSA analysts estimate the proportion in each group that paid taxes on at least \$5,000 in earnings, and sends me these group statistics – but only if the group is constructed in such a way as to meet SSA's confidentiality rules. By SSA rules, if all, or all but a few, persons in a group have no earnings, or have positive earnings, then the group must be combined with some other group until this is not the case.

I was concerned that some foreign nationals may not have SSNs but may report a nine-digit number that they use for identification instead, and also that some may simply make an error in recording their SSN. If so, my method could overestimate the stay rate if a U.S. citizen had the number they reported. To minimize this problem, the NORC sends the birth year associated with each SSN. Then the SSA follows a procedure that minimizes the possibility of a false match by treating those without a match on birth year as if they did not report their SSN to the SED.

The SSA sends me the proportion of each group paying taxes on at least \$5,000 in earnings in any specified year. SSA uses earnings subject to FICA tax to start with, but for persons paying FICA tax on less than \$5,000 in income, they also check Federal Income Tax records. For example, some postdoctoral fellows do not have to pay FICA tax because their award is classed as a fellowship, but in nearly all cases they do pay U.S. income tax¹. However, the proportion I receive from SSA could deviate slightly from the true stay rate for several reasons. One of those reasons is that persons might pay U.S. taxes even though they did not earn income in the United States. This is thought to be near zero, although there is no solid data to confirm that. Other reasons for a deviation between the true stay rate and the proportions with earnings reported to me by the SSA are taken into account. Adjustments are made as follows.

1. They might have died after receiving the doctorate. I use age-specific death rates to adjust for this, but the effect of such adjustment is very small.

¹ There are tax treaties between the United States and several foreign countries that can exempt a foreign national from paying U.S. income taxes under certain circumstances. Typically, they must work for a university or non-profit research institution. However, this is thought to have little practical impact as the exemption is limited to three years or less, and because many foreign nationals may exempt income earned before they complete their doctorate and thus exhaust their exemption even sooner.

2. They might have stayed in the United States but did not have earnings during the year in question, or earned less than \$5,000. I adjust for this by estimating the proportion of recently graduated, foreign-born S/E doctorates in the SDR who do not report earning at least \$5,000. Estimates of this proportion have a rather high standard error, and have varied considerably in the range of 0 to 4 percent depending on which cycle of the SDR is used to make the estimate. However, the average from four different SDR cycles has been in the range of 2 to 3 percent, and this is what I use for adjustments.
3. Those who report valid SSNs may differ from those for whom no valid SSN is available. We don't know just how the stay rate behavior of these two groups differs, so we must make some assumption. In the past I assumed that those who did not have SSNs stayed at only half the rate of those with SSNs; since the proportion missing SSNs was small, usually around 6 percent, this assumption was a way of finding a middle ground between the extremes of assuming that all those without SSNs left the U.S. and assuming that they were just as likely to stay as those with SSNs. I calculated that my stay rate estimates could not be off by more than 2-3 percentage points, and probably much less. Thus it did not make a great deal of difference.

In more recent years, however, there has been an increase to about 15% in the proportion of SED respondents who do not report SSNs. I do not have space to discuss how I am dealing with this, but will address the future of SSN-based stay rate estimates at the end of this paper.

Here are some of my recent findings:

- Stay rates increased dramatically during the 1990s but have leveled off in recent years. 2-yr stay rates have even declined slightly – but I think that is just a temporary response to post-9/11 events and will not continue. (Figure 1)
- Stay rates vary by field; agricultural and social sciences have the lowest stay rates.(Table 1)
- Stay rates decline only slightly from one year after the doctorate to 5 or 10 years after, even in fields where postdocs are common. (Table 1, Table 2)
- Doctorates from China, India and Eastern Europe have the highest stay rates (Table 3)
- Doctorates from Indonesia, Mexico, Brazil, South Korea, and Japan have the lowest stay rates
- The patterns described above have been very stable over the past decade; and will be largely unchanged in my next stay rate report.

Can this method be extended to estimate stay rates for foreign citizens receiving BS and MS degrees?

To date I have not attempted to use this method to estimate stay rates for foreign citizens receiving bachelors or masters degrees. Whereas the SED asks for SSN, the other NSF surveys do not. I know of no other adequate source of SSNs that could be used for this purpose. However, I recognize that there is an increasing need for this information. For example, I've received calls from Congressional staffers who were tasked with estimating the impact of bills that would provide increased access to H-1B visas for these graduates. If they don't know the current stay rate as a baseline, it's very difficult to even guesstimate the impact of changes in policy and law.

In light of this need, I discuss two methods, below, which could be used to estimate stay rates for bachelors and masters degree recipients. Each has flaws, but in each case I would propose to

use data from the doctorates stay rate estimates to minimize those flaws. For simplicity I discuss only the Masters recipients estimating method, but the Bachelors would be similar.

Estimate Using the Survey of Recent College Graduates.

I think the method I use to estimate stay rates for doctorate recipients is a good one. All I need to apply the same method to bachelors and masters recipients is a good source of Social Security numbers (SSN) for the bachelors and masters recipients. The Survey of Recent College Graduates (RCG) does not ask respondents to provide their SSN. However, when NSF asks the colleges and universities to provide lists of S/E graduates they also request SSN to help identify and locate the individuals selected for the survey. The problem is that some colleges and universities, roughly half, do not provide the SSN to NSF.

I know of no reason to suspect that that the universities which supply SSNs would have systematically different stay rates compared with the total of all foreign citizens on temporary visas who receive bachelors and masters degrees. However, to be credible we would need some way to have confidence that this is so.

I would like to develop a proposal to use the RCG to estimate stay rates in this manner. In light of the spotty cooperation of universities in providing SSNs for masters recipients, I would propose the following action.

I would estimate doctorate stay rates using the method I have used in the past. In addition, I would estimate the doctorate stay rate for the subset of universities that provided SSNs to the RCG. If the total doctorate stay rate is equal to, or very close to the doctorate stay rate for the subset of universities providing the SSNs for masters and bachelors recipients, then this provides a strong reason to assume that the masters and bachelors stay rates estimated from the RCG are representative of the national stay rate for all bachelors and masters.

Estimate Using the Post-censal survey together with tax-based data. This survey, known as the National Survey of College Graduates (NSCG) was conducted in 2003. It's coverage of Masters degree recipients during the 1990s is based on the self-response given by respondents to the 2000 Census of Population, and refined with the subsequent 2003 NSCG questions.

Conceptually, one might be tempted to use degree data to determine the number of U.S. Masters degrees awarded to foreign citizens during a given number of years, e.g., the 1990s, and then to use the NSCG to determine how many were still in the United States. This approach would encounter several serious problems. First, the degree holders identified in the NSCG are self-reported. Suppose that some of those who earned a masters "equivalent", e.g. two years of graduate study, reported an S/E masters when their university did not actually award them the credential. This would lead to an overstatement of the stay rate. Also, persons with degrees closely related to an S/E field might self-report in an S/E degree field, or visa versa. We know from studying the 1993 NSCG that the NSCG produced higher counts of S/E doctorates than did the Survey of Doctorate Recipients. (Finn and Baker, 1995) The same is probably also true of the 2003 NSCG.

Another problem with using the NSCG to estimate stay rates is non-response bias. Foreign degree recipients are more likely to return the survey if they are in the United States than if they are not. However, the non-respondents are assumed to be like the respondents who have similar characteristics – so an overestimate of foreign S/Es in the United States typically occurs in surveys of this type.

A third problem with this approach is that the degree data available for Masters recipients does not identify foreigners with permanent resident visas -- they are included with U.S. citizens.

However, the NSCG survey does not permit us to identify type of visa held at the time a degree was received.

In spite of all these problems I tried, for this meeting, to use the NSCG together with degree data and my estimates of doctorate stay rates to produce estimates of stay rates. I start with the number of masters and doctorate degrees awarded to foreign citizens on temporary visas during the period from 1989 to 1998.

During 1989 to 1998 U.S. universities awarded 2.6 masters for every doctorate awarded to foreign citizens on temporary visas.

If the masters stay rate were equal to the doctorate stay rate we would expect to see 2.6 times as many masters recipients as doctorate recipients from this era in the NSCG.

Do we see that many? No, we have only 76 percent of the number we would expect if the stay rates were the same.

However, that is misleading because some foreign-born S/Es earn both a Masters and a Doctorate from a U.S. university. Those who earned both are counted as doctorates in the total for both the masters and the doctorate categories in the degree data, but counted only as doctorates in the NSCG data – because I tabulated by highest degree held. Could we identify those with both degrees and make the appropriate adjustment. Yes, but I have not done that yet.

Let's explore this method, however, by assuming that some foreign-born doctorate recipients also received masters degrees from U.S. universities.

These data suggest that the masters stay rate during the 1990s would be equal to the doctorate stay rate if we assume that 60 percent of the doctorates also earned U.S. masters degrees. It would be 87 percent of the doctorate stay rate if we assumed that only 25 percent of the doctorate recipients earned U.S. masters degrees. The SDR would allow us to estimate the actual number and my guess is that it would be between these extremes, but closer to the lower one.

I see these problems with this method of estimating masters stay rates:

- 1.) We can only do it once each decade.
- 2.) It would be biased if the inclusion of degrees awarded to foreigners with permanent resident visas would have changed the doctorate/masters degree ratio of 2.6. This can be checked and, if needed, adjustments could be made. The SED tells us how many doctorates were awarded to permanent resident visa holders, and the NSRCG can do the same for masters degree recipients.
- 3.) There is another critical assumption that may not hold true. In the case of doctorates, the number of foreign-born doctorates recorded in the NSCG survey was 15 percent higher than the number obtained by applying my stay rate estimates to the number of doctorates awarded to foreign and naturalized citizens recorded by the SED. There are some good and some not-so-good reasons why the NSCG gives higher number of doctorates than the SED. Good: the SED is restricted to "research doctorates" and the NSCG isn't, so, for example, a person with a PsyD or other practice-oriented science doctorate is included in the definition of the NSCG but not the SED. Not-so-good: some people exaggerate their credentials when responding to a survey, but this is not possible with the SED as the university certifies that these are doctorate recipients. If we estimate masters stay rates using the NSRCG and the method described above we have to assume that the factors which cause the number of foreign-born doctorates in the NSCG to exceed those in the SED apply in a proportional manner for the masters degree recipients.

However, the fact that the number of foreign-born doctorates in the NSCG was only 15 percent higher than the number that could be obtained by combining my stay rate estimates with degree data suggests that the problem of overestimation in the NSCG is very great. This method could

be used to produce estimates for the those graduating during the 1990s relatively quickly and easily.

Difficulties with the use of Social Security numbers for estimating stay rates.

I've already mentioned the fact that increasing numbers of respondents to the Survey of Earned Doctorates do not supply their SSNs. However, foreign nationals are no more likely to withhold SSNs than U.S. citizens. If we compare this method of estimating stay rates with using a survey such as the SDR to determine stay rates, the survey still seems to be inferior. Even the best surveys have comparable problems with non-response, but non-response bias is arguably more of a problem with a survey based estimate of stay rates. Non-response bias studies have shown that, in the past, the SDR overestimates the presence of foreign nationals more so than other demographic groups. The reason is that people who leave the United States are less likely to receive and/or return the survey, and the methodology assumes that the non-respondents are like the respondents in each survey stratum. Foreign nationals are much more likely to leave the United States than other demographic groups.

NSF has responded to reticence to supply SSNs. In the future they will not ask for SSNs. At least in some cases I understand they will ask for the last four digits of the SSN, e.g., when collecting student information from colleges and universities participating in the NSRCG. This will not make it impossible to use the Social Security Master Earnings file to track graduates, as one can match on name and birth date. However, it will make it much more time-consuming and expensive. I have been sending tens of thousands of SSNs to SSA for tabulations. Cost and time constraints mean this method would probably need to be used with smaller groups in the future.

However, in this regard, the future is a good way off. I will soon release the 10 and 5 year stay rates for the classes of 1995 and 2000. The fact that complete SSNs will not be available for the doctorates of 2008 will not affect the release of 5 year stay rates using this method for another eight years.

References

Finn, Michael G., and Joe G. Baker, "Reconciling Estimates of U.S. Doctorates", Oak Ridge, TN: Oak Ridge Associated Universities, November 1995

Finn, Michael G., "Stay Rates of Foreign Doctorate Recipients from U.S. Universities, 2003," Oak Ridge:TN: Oak Ridge Associated Universities, November 2005.
<http://orise.orau.gov/sep/files/stayrate05.pdf>

Mitchell, Susan and Daniel Pasquini, (1991) "Nonresponse Bias in the 1989 Survey of Doctorate Recipients: An Exploratory Study," Office of Scientific and Engineering Personnel, Washington, D.C.: National Research Council

Moonensinghe, Ramal, Susan Mitchell, and Daniel Pasquini, (1995) "An Identification Study of Nonrespondents to the 1993 Survey of Doctorate Recipients", in Proceedings of the Section of Survey Methods. Alexandria, VA: American Statistical Association.
<http://srsstats.sbe.nsf.gov/docs/research/3-51.pdf>