Should I Stay or Should I Go ... North?

First Job Location of U.S. Trained Doctorates 1957-2005

Christopher Ferrall [†]   Natalia Mishagina [‡]

Department of Economics
Queen's University

September 10, 2007

Abstract:

This is an extended abstract / preliminary draft paper submitted in response to the Call for Papers for the NBER Universities' Research Conference on The Economics Of High-Skill Labor Markets, December 2007. All empirical results planned for the full paper have been carried out. This document presents only the baseline specification in order to demonstrate that a number of novel and statistically significant results about international mobility of the highly skilled have been found.

Correspondence:
[†] Associate Professor and planned presenter of the paper (ferrallc@queensu.ca).

[‡] PhD Candidate.

I. Motivation

The production of human capital is a global phenomenon, especially within post-graduate university education. International movements of skilled workers seems is closely related to movements of workers that seek training. For example, in 2005 the U.S. granted over one million visas to students and their families, about 200 thousand more than the visas issued to temporary workers and their families. About one half of the worker visas were to specialty occupations (H1B). Undoubtedly a large number of those people were moving from student status to temporary worker status.[1] This process is also seen first hand within academic departments each year with the arrival of international graduate students without definite plans who ultimately immigrate permanently to the host country.

This paper concerns the first jobs after receiving a doctorate degree from a U.S. university. The flow of people in search of skill acquisition plays an important role in the post-training allocation of that skill. What determines the location of newly minted PhDs? Is the decision to stay in the U.S. related to the background of the student? How important is the relative economic situations of the U.S. and the home country? Do other policies or international situations matter?

Answers to these questions determine which economies benefit from the human capital imparted by specialized training. The answers also influence whether governments that support direct cost of educating their citizens (or others) capture the benefits. New PhDs are usually minted at a point in the lifecycle with a high degree of mobility, both geographic and economic. The credential is universally recognized as representing skills at the leading edge of knowledge. Through their teaching and research activities, the location decisions of new PhD may help determine international growth rate differences in future decades.

---

[1] Figures taken from Table 25 of Yearbook of Immigration Statistics: 2005.

II. Focus of this Paper

Our story begins at the point of earning a PhD from a U.S. university because data are limited on who decides to study abroad and why. The proportion of the populations that are both highly skilled and liable to international mobility is small. And the process for selecting students is decentralized. Thus, data on who moves to do a PhD are not likely to be available at the global level. However, since U.S.-based post-graduate education collects and concentrates a large fraction of the worldwide population of the highly able, it is possible to observe the location decision after the degree. Fortunately the unique role of American post-graduate education is captured by an equally unique but under-used source: the National Science Foundation's Survey of Earned Doctorates (SED). Despite its name, the SED is a year census of all PhDs earned in the US since 1957. The SED is provided to graduates by their university while completing the final requirement for their degree. This means the coverage is excellent and the timing of the survey is such that the preponderance of the respondents have definite if not already executed plans for their first post-doctorate job.

This wealth of information still suffers from two problems. First, only a small fraction of American graduates take their first position outside the U.S. Second, international students are coming from widely varying situations that are difficult to measure and many will simply return home. To circumvent these issues, we study the location of first jobs between the U.S. and Canada conditional on reporting one of the two countries as the destination. Although only a small fraction of all U.S.-educated doctorates report Canada as their destination, the census aspect of the SED makes the sample size large enough that significant effects are easily obtained. And tightening the focus to U.S. versus Canada makes it straightforward to include the difference between unemployment rates as a measure of prevailing relative labor market conditions.[2] We compare the

---

[2] In contrast, while other countries could be included, a quickly vanishingly fraction of doctorates would go to other countries except for citizens of that country.

propensity to locate to Canada among U.S.-educated Americans, Canadians, and third-country nationals (3rdCNs). This means that we are able to observe home-bias that works in opposite directions and is neutral.

Besides controlling for many standard demographic factors entering the location decision of graduates, there are two other revealing variations to study inside the SED. First, both economic and non-economic policies have been important over the sample period. The sample for the basic set of variables extends from 1957 to 2005 which brackets the Vietnam War era. It also includes the era of reduced barriers to labor markets brought about by the Canada-U.S. and North American Free Trade Agreements. The our sample also extends far enough past the events of September 11, 2001 to study the choices of 3rdCNs who chose the U.S. before that date finished their degrees after.

The second set of deeper determinants that we can study come from the SED survey. In some periods within 1957-2005 the SED asked recipients about their parents education, whether they had school-related debt at graduation, and whether they received foreign funding. Since the funding issues are closely related to public policies designed to enhance the (domestic) supply of skill, the SED provides a unique opportunity to see how such factors help determine international mobility and the so-called brain drain.

III. Empirical Framework

We start by considering a general model for new graduate i planning to move to destination country d having latent utility $y_{id}^{\star}$. The number of destinations is large and the number of non-natives graduates observed moving to most destinations is very small. Construction of destination-specific controls such as the current unemployment rate is difficult to implement at a global level.[3] For these reasons we restrict the sample and the

---

[3] Even if the first transition after the PhD were not a singular career event, the SED's design nature allows only a cross-sectional analysis. The NSF does indeed conduct a panel survey on a restricted subsample of the SED universe, but it cannot be used to

model to the case where d is either Canada or the U.S.

We hypothesize that movement from the U.S. to Canada after completing the degree depends on individual preferences, relative demand for specialized training between the two countries, and policies that encourage or discourage the individual to move. Setting the U.S. as the default choice only the conditional latent index for Canada remains. A probit model that nests all our specifications takes the form:

$$y_{io}^{\star} = f_o\left(DEMO_i, PGM_i, UNIV_i, FIN_i, y_i, POLICY_{y(i)}, \Delta U_{y(i)}\right).$$

Here o denotes the origin of graduate i, which takes on the values US, CDN, or $OTHER$ (3rdCN). The function f is indexed by o because we estimate separate probits by origin. As one would expect the hypothesis that the three origins share coefficients on the included variables is decisively rejected by the data. The arguments to f_o are two types of vectors. First are vectors of characteristics specific to graduate i as measured in the SED. We group them into four vectors: DEMO contains demographic background variables; PGM contains indicators for the graduate's field of study; UNIV contains indicators for the quality ranking of the graduate's university; and FIN contains variables related to the graduate's financial arrangements. The second set of variables depend on the year of graduation denoted y_i. Beside including y_i itself (and considering a quadratic time trend in some specifications), we include two sets of year-specific controls in some specifications. POLICY includes indicators for whether various policies that would affect the choice between Canada and the U.S. are active in year y; ΔU is the relative annual unemployment rate between Canada and the U.S. in the year of graduation:

$$\Delta_y = \text{US UE Rate}_y - \text{Canada UE Rate}_y.$$

Obviously there are both demand and supply factors determining the value of $y_{io}^{\star}$. In some cases, for example when looking across fields indicated by the PGM vector, we

---

detect determinants of international flows in subsequent job changes. Mishagina (2007a and 2007b) analyze the career dynamics of scientists and engineers using that data set.

can only observe the net effect of the two sides of the market. On the other hand, as we discuss below, some supply and some demand concerns affect individuals differently by their country of origin.

IV. Features of the Data

The overall sample size is 1,603,141 observations. Of those, 84.4% respondents state they have post-degree 'plans'. The figure below shows the nature of the plans by citizenship, indicating that most plans are very concrete. Even among those that are still seeking jobs, the number that would ultimately end up in a different country that the planned one would likely be quite small. Thus, the signal on planned location asked several months after completing a doctorate is likely to contain a great deal of information on ultimate outcomes.

Table 1 describes the variables contained in the four individual-specific vectors DEMO, PGM, UNIV and FIN. The richness of the data and our restriction to the U.S.-Canada choice allows us to isolate several forces that would affect individuals differently depending on origin. For example, the SED asked for parental education for forty years. We use the traditional interpretation of this as an indicator of ability (controlling for field and university). We posit that mobility is positively correlated with talent. However, we mean mobility relative to any 'home' bias. Thus, parental education would tend to have a negative effect of returning to Canada for Canadians but a positive effect for Americans and 3rdCNs. Meanwhile, graduates with dependents are hypothesized to be less mobile and the signs for this variable would be the opposite of those for parental education. We can also consider that language is an important determinant of location and Canada is a bilingual country. The SED asks whether English is the graduates native tongue. But we go further and categorize all graduates by whether they attended high school in a Francophone country. These graduates are hypothesized to be more likely to move to

Canada, all else constant. While this is a straightforward prediction, the ability to check it and similar predictions against the results increases confidence in the interpretation of other factors within the mobility model.

The limited spans for some of the variables justify the reporting of different specifications. For example, we have already suggested that including parents' education is a useful (and conventional) control for ability. However, including in the specification precludes a study of the post 9/11 outcomes since the questions were dropped in 2002. Similarly, the short and late coverage of the financing questions preclude a study of the Vietnam Draft and NAFTA eras.

V. Baseline Model

Table 2 reports selected coefficients from the baseline probit model (explaining whether planning to locate in Canada conditional on plans for Canada or the U.S.) It includes only variables available for the full span of 1957-2005. Also included in the model are the PGM and UNIV vectors of controls. Standard errors are reported and coefficients that are significantly different from zero at the 5% level are indicated. It can be seen from the table that many of the predictions from a simple mobility model are consistent with the data. For example, temporary visa students are more likely to move to Canada as are American and international students who went to high school in Francophone countries. Married students and female students have a home bias (American coefficient is negative, Canadian positive). Conditional on staying in the U.S. and Canada, both male and married 3rdCNs are more likely to move to Canada.

Years in which the U.S. - Canada unemployment rate is greater leads Americans and 3rdCNs to be more likely to move to Canada. One might hypothesize that such a coarse statistic from the labor market would have low correlation with the small and specialized market for PhDs. And it might be the case that the contemporary conditions would have

little effect on job location when making what is often a lifelong decision. However, the effect is large and significant for both Americans and 3rdCN graduates. The sign is the opposite for Canadians but insignificant. We expect third-country nationals to be most sensitive to this difference since neither the U.S. or Canada is home and the UE Rt difference may be correlated with other factors affecting the home bias. Indeed, the effect of the UE Rt is about 75% larger for 3rdCNs than Americans.

During the Vietnam War graduates of each origin had a higher propensity to move to Canada, although the effect is not significant for Canadians. (Perhaps this is because Canadians could easily work in the U.S. without becoming eligible for the draft than either Americans or 3rdCNs.) During the NAFTA period Canadians are less likely to return home, and unexpectedly Americans are less likely to move to Canada. For third-country nationals the effect is small and insignificant, which is consistent with the mobility model since NAFTA had no direct impact on either the demand or supply side of their situation. NAFTA should encourage Americans to move to Canada, so the negative coefficient suggests the year dummies are picking up effects not captured by the linear trend. But consider that the impact of NAFTA was not symmetric by country of origin. For Canadians it lowered barriers to employment in an economy roughly ten times the size of their domestic market. And Canadians in the SED have already lived in the U.S. for several years and have close contacts and a familiar pedigree. For Americans NAFTA lowered barriers to a market of about the size of California, but a market with which many would have no prior contacts. Thus, the mobility effect of NAFTA should be much larger on Canadians than Americans. Indeed the estimated propensity is larger for them in absolute value. So the differential impact of NAFTA may be captured by the difference in the two coefficients (although this is yet to be worked out).

Finally, there is evidence in the baseline model that 3rdCN graduates were more likely to move to Canada after 9/11/2001. Nearly all of the graduates in our sample would have chosen the U.S. as location to study before 9/11. Except for students who left before

graduation due to 9/11, the SED captures the relative preference for the U.S. and Canada alone without any response in the population choosing to study in the U.S. after 9/11. The negative time trend for 3rdCns indicates less propensity to move to Canada over time. And while the 9/11 coefficient is only a fraction of the size of the temporary visa and non-English coefficients, the 9/11 effect is over 10 times the size of one year of their time trend. Thus international mobility of the highly skilled appears sensitive to political events.

We conclude that aggregate economic conditions (ΔU), specific economic policies (NAFTA) and non-economic policies and conditions (Vietnam, 9/11) all have significant effects on the group of graduates that are most likely to be affected by them. This remains even after controlling for a large number of other factors that strongly influence mobility (including field, university quality, and personal characteristics).

VI. Further Analysis

The completed paper presents results from models that include other variables not available throughout the SED time span. It also analyzes the sensitivity of key results to the specification. The rich combination of mobility factors and the comparison of the coefficients across citizenship will be more fully worked out to determine the sign of some of the purified effects of the time factors.

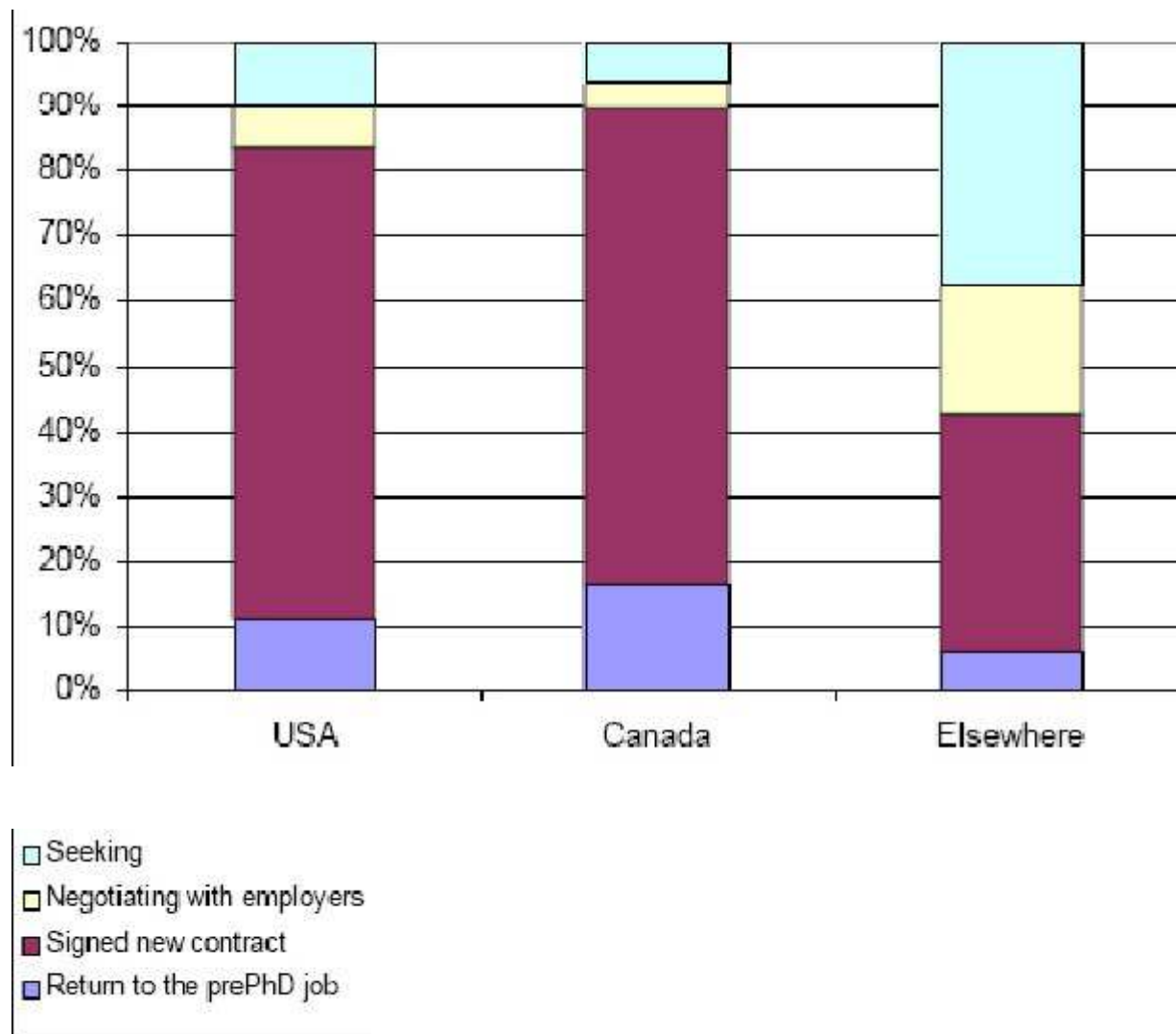# Figure 1. Post-Doctorate Plans By Citizenship

Seeking

Negotiating with employers

Signed new contract

Return to the prePhD job

Table 1. Summary Statistics

Variable	Span	Overall Obs.	Overall Mean	Mean By Plans USA	Mean By Plans Canada	Mean By Citizenship Amer.	Mean By Citizenship Cdn.	Mean By Citizenship Other
Demographics								
Temporary visa	57----05	1603141	0.158	0.113	0.608	0.000	0.754	0.489
Non-English (1st language)		1603141	0.338	0.196	0.237	0.049	0.144	0.971
Francophone***		1603141	0.022	0.016	0.462	0.005	0.882	0.025
Married		1592281	0.578	0.665	0.654	0.651	0.606	0.413
Male		1600951	0.699	0.672	0.776	0.657	0.734	0.795
Father college degree	62----02	1278493	0.320	0.322	0.284	0.318	0.246	0.330
post-graduate work		1274706	0.222	0.230	0.206	0.234	0.205	0.187
Mother college degree		1273531	0.332	0.344	0.303	0.362	0.287	0.230
post-graduate work		1272086	0.111	0.118	0.091	0.121	0.087	0.077
Has dependents	72----02	1139026	0.548	0.544	0.562	0.540	0.483	0.583
Asian		1103755	0.172	0.145	0.144	0.030	0.081	0.607
Black		1103747	0.048	0.045	0.012	0.047	0.013	0.052
Hispanic		1116056	0.037	0.033	0.014	0.029	0.005	0.063
Funding								
Has school debt	82----02	498567	0.484	0.500	0.473	0.559	0.494	0.297
PhD foreign-funded	86----02	900763	0.024	0.015	0.057	0.004	0.104	0.095
University Type*	**57----05**	1603141						
Topschools & Carnegie 1			0.132	0.113	0.197	0.110	0.238	0.175
Carnegie Category 1			0.892	0.881	0.926	0.882	0.913	0.918
Carnegie Category 2			0.076	0.084	0.050	0.084	0.054	0.056
Field of Study**	**57----05**	1603141						
Computer sciences			0.011	0.013	0.010	0.008	0.010	0.018
Mathematical sciences			0.029	0.027	0.043	0.024	0.035	0.039
Life and Health sciences			0.182	0.184	0.192	0.175	0.196	0.195
Physical sciences			0.128	0.121	0.123	0.111	0.110	0.166
Social sciences			0.083	0.089	0.064	0.103	0.076	0.038
Psychology			0.087	0.079	0.134	0.081	0.113	0.098
Engineering			0.112	0.113	0.070	0.076	0.083	0.193
Humanities and Law			0.318	0.320	0.283	0.370	0.306	0.205

**Default is Business Administration

***Person finished high school in a country where French is an official language

*Category 2: School awarded at least 10 doctoral degrees per year across 3 or more disciplines or at least 20 doctoral degrees per year overall.

Category 1: School awarded 50 or more doctoral degrees per year across at least 15 disciplines.

Top School: Category 1 AND CalTech, UC Berkeley, Yale, Harvard, MIT, Columbia, Cornell, Stanford, or Princeton.

Table 2. Baseline Probit Model

Variable	American	Canadian	Other
non-english mother tongue	-0.070	-0.124 *	-0.465 *
	(0.04)	(0.04)	(0.03)
francophone	1.225 *	-0.032	0.412 *
	(0.04)	(0.04)	(0.03)
temporary visa		1.503 *	0.750 *
		(0.03)	(0.02)
married	-0.067 *	0.325 *	0.029 *
	(0.01)	(0.02)	(0.02)
male	0.100 *	-0.105 *	0.075 *
	(0.01)	(0.03)	(0.02)
year of graduation	0.003 *	-0.022 *	-0.011 *
	(0.00)	(0.00)	(0.00)
UE RT Diff	0.030 *	-0.013	0.052 *
	(0.00)	(0.01)	(0.01)
NAFTA	-0.151 *	-0.238 *	0.027
	(0.02)	(0.04)	(0.03)
Vietnam war	0.263 *	0.011	0.252 *
	(0.01)	(0.03)	(0.03)
Post 9-11			0.129 *
			(0.07)
N	850856	16175	170401
ll	-32415	-8655	-15150
chi2	3238	5103	3492