ESSAYS ON THE CAREER CHOICES OF DOCTORAL STUDENTS IN THE U.S.

by

Chandramouli Banerjee

A Dissertation Submitted in

Partial Fulfillment of the

Requirements for the Degree of

Doctor of Philosophy

in Economics

at

University of Wisconsin–Milwaukee

August 2018

ABSTRACT

ESSAYS ON THE CAREER CHOICES OF DOCTORAL STUDENTS IN THE U.S.

by

Chandramouli Banerjee

The University of Wisconsin – Milwaukee, 2018
Under the supervision of Professor Niloy Bose

This thesis consists of three essays on the post-graduation career choices of doctoral students in the U.S. and the impact these choices may have on innovation and the competitiveness that the U.S. enjoys in the global science and engineering landscape. The first chapter studies the location choice of work of foreign-born U.S. doctorates, who have been playing a central role in shaping the U.S. skilled workforce over the past few decades. Evidence suggests that not all foreign-born U.S. doctorates choose to remain in the U.S. following graduation. This chapter uses a new data set - the International Survey of Doctoral Recipients (ISDR) - to identify a number of demographic and country specific factors having implications for location choice of work for foreign-born U.S. PhDs. In addition, we find evidence of a temporal increase in the intensity of positive skill selection among foreign-born U.S. PhDs leaving the U.S. workforce. The result indicates that U.S. may be losing premium talent to global competition.

The second chapter studies the choice of the type of job that a S&E doctoral student matches with and how job-skill match in the labor market for scientists impacts productivity at the industry level and hence innovative processes at the

aggregate level. This chapter primarily offers a transparent theoretical approach that demands relatively little from the data and yet produces reliable estimates of the output gain due to job-skill match in the labor market. We apply this approach to data containing information on job choices of scientists in the U.S. The results suggest that for all major skill types/industries, job-skill match creates larger value as opposed to skill mismatch. At the same time, the estimated match surplus responds differently to economic conditions across industries. This difference is useful for uncovering important industry specific traits, including an industry's propensity toward diversification and innovation. In addition, we investigate the relationship between the output gained due to a skill match and innovation at an aggregate level. We find that an increase in a market index of output surplus generated by the skill match increases research output in the economy, as measured by total patent applications. This points to a channel through which the effects of job-skill match could show up in the form of higher productivity.

The third chapter builds on the findings in the first chapter by attempting to uncover the causal relationship between attending a highly ranked graduate program in the U.S. and the propensity to leave following graduation for foreign-born U.S. doctoral students. A variety of unobservable factors at the individual level that may affect the attendance in top programs and propensity to emigrate may attenuate the correlation that is picked up in naive OLS regressions. To isolate the effect of attending a top program on the probability of leaving we instrument top program attendance at the individual level by the average past top program attendance from the students' country of origin. The instrument is plausibly correlated with top

program attendance since a greater number students attending top programs from a particular country may encourage others from that country to apply to these programs. Additionally, this may induce top programs to admit more students from a particular country since these programs have better information about the quality of education in the country of origin through past students. The IV results, while confirm the findings of the first chapter, also find that the naive OLS regressions underestimate the impact of top program attendance on probability of leaving the U.S. following graduation substantially.

Thesis Supervisor: Niloy Bose

Title: Professor of Economics

To Ma, Baba and Nikita.

TABLE OF CONTENTS

Table of Contents vi

List of Figures viii

List of Tables ix

Acknowledgments xi

Introduction 1

1 To Stay or Not to Stay: Location Choice of Foreign Born U.S. Doctorates 11

1.1 Introduction 12

1.2 Related Literature and Challenges 16

1.3 Data and Variables 19

1.4 Model and Results 25

1.5 Concluding Remarks 33

2 Job-Skill Match in the Labor Market for Scientists and Its Aggregate Implications 45

2.1 Introduction 46

2.2 A Model of Worker-Firm Matching 50

2.3 Empirical Results 57

2.4 Conclusion 68

3 Skill Selection in Out-Migration of Foreign Born U.S. Doctorates: A Causal Approach 78

3.1 Introduction 79

3.2 Constructing the Instrument 82

3.3 Model and Results 85

3.4 Concluding Remarks 89

3.5 Tables 91

References 95

A Appendix A: Variable Names and Data Sources (Chapter 1)101

B Appendix B: Degree Field Groupings (Chapter 2)103

C Appendix C: Job Category Classifications (Chapter 2)108

Curriculum Vita 112

LIST OF FIGURES

1.1 U.S. research doctorates awarded in science, engineering, or health, by citizenship: 1960 - 2010 37
1.2 Trend in Doctorates awarded by Top Programs 37
1.3 Trend in Leave Rates of Top Program Students 38
1.4 Intensity of Positive Skill Selection by Decade 1970-2011 38

2.1 Added Value Generated by a Skill Match by Subject Area over Time . . 71
2.2 Added Value Generated by a Skill Match by Subject Area over Time (Contd.) 72
2.3 Market Index of Added Value Generated by a Skill Match 73
2.4 Impulse Responses from VAR Analysis 73

LIST OF TABLES

1.1 Foreign-born workers in S&E Occupations, by Education Level: Selected Years, 2000 - 11 39

1.2 Descriptive Statistics 40

1.3 Determinants of Leaving the U.S. Following Receipt of a PhD 41

1.4 Determinants of Leaving the U.S. by country of Country of Origin and Job Type 42

1.5 Estimates for Students from India, China, South Korea, Taiwan: Incorporating Match Quality and Salary Premium 43

1.6 Estimates for Students from Low/Middle Income Countries: Where do they go? 44

2.1 Added Value Generated by a Skill Match: Summary Statistics 74

2.2 Correlations of Surplus Value with Lagged Two Year Average of GDP Growth 74

2.3 Correlations with Lagged Two Year Average of GDP Growth: Certain Fields over Time 75

2.4 Parameter Estimates from Dynamic Factor Analysis: Factor Loadings . . 75

2.5 Parameter Estimates from Dynamic Factor Analysis: AR(1) Parameters 76

2.6 Parameter Estimates from Dynamic Factor Analysis: Innovation Variances 76

2.7 Proprtion of Variance Explained (PVE) by Common Factor 77

2.8 VAR Results 77

3.1 First Stage Regressions: Full Sample and S&E Only 91

3.2 Determinants of Leaving the U.S: OLS vs. IV for Full Sample and S&E Only 92

3.3 Determinants of Leaving the U.S. by Country of Origin: OLS vs. IV . . 93

3.4 Determinants of Leaving the U.S. by Country of Origin for S&E Only: OLS vs. IV 94

A.1 Variable Descriptions and Data Sources 102

C.1 Job Category Classifications 109

ACKNOWLEDGMENTS

Having been fortunate enough to be supported and guided by the most amazing humans through my life, I have rarely felt the need to lean on the divine. This little prose, finally, gives me a brief opportunity to thank all those who have helped lead me up to and then see me though the journey of my doctoral career. Let us hope that I do this justice.

First of all, I am extremely grateful to my thesis supervisor Dr. Niloy Bose, for being more of a father than a mere mentor through this process. This dissertation was shaped through constant debate and consistent collaboration with Dr. Bose over the past four years. He pushed me to explore the boundaries of my intellect, guided me when I wavered, and most importantly believed in me enough to give me his time. I have learned much more than Economics under his tutelage and however the canvas of my life looks like going forward, his brushstrokes will always be visible. For this, I am forever indebted.

I would like to thank the other members of my dissertation committee, Dr. Scott Adams, Dr. John Heywood and Dr. Kundan Kishor for their helpful comments, career advice and in no small measure, for listening patiently while I rambled on about my work. To be believed in during a very uncertain time of one's life is worth more than I can express and I am grateful to these gentlemen for doing so. I would be remiss if I didn't also thank Ms. Brenda Cullin, the matriarch of the department, who made sure I was fed and clothed and fought the administrative fights on my behalf while I focused on working.

Next, I would like to thank those who guided me onto the path of pursuing a doctoral degree. Ms. Sandhya Raman and Mr. Robin Colaco, without whose urging and care I would have been mathematics illiterate; Ms. Archana Agarwal whose undergraduate class made me fall in love with Economics and Dr. Niti Bhutani, whose unwavering belief in my research capabilities gave me the confidence to take off running on this path. I would like to thank my friends through times before, during and presumably beyond graduate school who have all played some part in my time therein. Shomik Mullick, Pratheek Vinod Kumar, Sarath V.M., Zoish Tosher Hormusjee and Nitish Kallianpur brothers, who have been and who I know will always be there; Tanvi Girotra, Kaavya Gupta and Shashvat Sikroria, for letting me be the 'nerd' of the group in college and letting me believe in that role; Adil Amin, Satyanshu Kumar, Deep Chatterjee, Shrathinth Venkatesh, Eduard Storm, Loren Peter Wagner, Amid Ardakani and Mrinmayi Kulkarni for accompanying me on the often arduous journey through graduate school; and Sahar Milani, Eric Shinnick and Haniyeh Arabi for making Wisconsin feel warmer than it actually is. If I've forgotten anyone, please know it is not for the lack of trying to remember.

Finally, I'd like to thank those who form the pillars of who I am as a person. My parents, for having worked so hard to give me any opportunity at being better. My father, for his smile, warmth and undeniable positivity, who manages to light up any day that seems dark. From my first examination to the very last, he has stood by me, in spirit if not in person, squeezed my hand just a little before I went in and I truly believe that this has tipped the balance in my favor many times. My mother, for giving the fire and the drive to be better today than I was yesterday, for holding us as a family together, for being brutally honest when I needed it and then staying up all night with me to make sure I succeed. Last, but not in the least, Nikita, the source of my joy and inspiration, for being by my side over the last six years while I traveled to a faraway land to embark on this journey. She has championed me at every step, loved me unconditionally at my best and my worst, mended me during the times I felt broken and put a smile on my face one way or another, every single day. I thank her for being in my corner, cheering me on. This would not have been possible without her.

INTRODUCTION

The impact of the very highly skilled members of the workforce on the sustained well-being of an economy through the creation and adoption of scientific knowledge is well established in the economics literature. The skills that these individuals own are key inputs in innovative processes in any economy, which in turn spur job creation and long term economic growth. Scientists and engineers are thus believed to be essential to technological leadership, innovation, manufacturing, and services, and thus vital to economic strength and societal needs of an economy. There is evidence to suggest that the skills owned by the S&E workforce are highly demanded by markets – doctoral students in S&E occupations enjoy very low levels of unemployment, are highly employable over time and contribute significantly to the knowledge economy. There is also evidence that this demand for S&E skills has intensified over the past few decades and will continue to grow in the future. In the face of the demand for S&E skills, the U.S. government has enacted many programs to ensure an adequate supply of these skills in the workforce through education and development of scientists and engineers by providing funds that encourage graduate and postgraduate research at U.S. colleges and universities through the financing of university-based research. Additionally, policymakers have sought to increase the number of foreign scientists and engineers working in the U.S. through various immigration programs that invite foreign scientists to study, work and innovate in the U.S. Such policy measures are not limited to the U.S. only – in an increasingly globalized economy, many other countries who recognize the value of having a highly skilled S&E workforce have been rapidly tailoring their policies to foster and attract talent. Given the central role that doctoral students play in the global knowledge economy, the career choices of these individuals with respect to where they choose to work, whom they choose to work with and how productive they are as result have a crucial impact on innovative processes in the U.S. and in the long run, the global leadership that the U.S. enjoys in research and development. This thesis studies the post-graduation career choices of doctoral students in the U.S. - in terms of location choice of work and matching behavior of doctoral students with different firms by leveraging detailed data on doctoral students in the U.S.[1] It explores the factors that affect these choices, and the implications that they have on U.S. innovation and research productivity.

The first chapter of this dissertation aims to learn about the location choice of work for foreign-born U.S. graduates. For the specific case of the U.S., the growth in the demand for highly skilled workers is partly satisfied by an increased participation of foreign-born scientists in the U.S. S&E workforce. Currently, foreign-born scientists represent a large fraction of the S&E workforce.[2] Significantly, amongst the foreign-born doctorate holders employed in the S&E workforce in the U.S., many received their doctoral degree from a U.S. institution and this supply pool of U.S. trained foreign-born doctorates has grown rapidly over the last few decades. The share of foreign nationals earning doctorates in Science, Engineering or Health in the United States was about 17% during the decade of 1960. By 2010, this share increased to

[1]Survey of Earned Doctorates (SED), Survey of Doctorate Recipients (SDR) and International Survey of Doctorate Recipients (ISDR) all complied by the National Science Foundation (NSF).

[2]According to the American Community Survey (ACS) and the Scientists and Engineers Statistical Data System (SESTAT), 26-27% of respondents employed in S&E occupations during 2010 were foreign born and the corresponding number for the pool of respondents with a doctoral degree is about 42-44%.

nearly 40%.[3] There is evidence to suggest that this effect is being driven by a large influx of students from low and middle income countries Grogger and Hanson (2015). The share of doctorates awarded to foreign-born students in the top tier universities has also grown rapidly while the total share of doctorates handed out by these schools has remained fairly constant. Foreign-born students now also dominate the pool of PhD recipients in many key subject areas.[4] The above stylized facts indicate that over the past few decades, foreign-born U.S. PhDs have shaped and are continuing to shape the landscape of the U.S. S&E higher education and workforce. It is, therefore, crucial to have a deeper understanding of the migration behavior of this group of individuals.

This goal is meaningful for a variety of reasons. First, there is evidence that a large fraction of foreign born S&E graduates emigrate after graduation. For example, Finn (2010, 2014) constructs stay rates of foreign-born doctoral students in the U.S. using Social Security Data and finds that approximately two-thirds of foreign-born doctoral students in the U.S. leave within two years of graduation. Second, there is compelling evidence that U.S. trained foreign born graduates make significant contribution to research and innovation. According to Chellaraj et al. (2008), a 10% increase in the size of foreign graduate students in Science and Engineering (S&E) fields leads to 4.5% increase in the university patent applications and 6.8% increase in the university patent grants. Similarly, the estimates by Stuen et al. (2012) suggest

[3]Source: InfoBrief, National Center for Science and Engineering Statistics, NSF-13-300, October 2012.

[4]Bound et al. (2009) report that students from outside the US accounted for 51% of PhD recipients in S&E in 2003, up from 27% in 1973. This trend holds across fields. For example, the same study finds that in 2003, foreign-born individuals accounted for: 50% of degrees in Physical Sciences, 67% of degrees in Engineering, 68% of degrees in Economics.

that having an additional foreign graduate student in S&E departments translates into an average gain of 5 extra articles in the department over the course of a doctoral student's 6-year graduate career. Hunt and Gauthier-Loiselle (2008) suggest that foreign-born students not only do not crowd out natives from the graduate school, but a one percentage point increase in the share of immigrant college graduates in the population increases patent per capita by about 15%. Beyond the domain of graduate schools, the contributions of the foreign-born graduates are disproportionately large as well. According to Peri (2007), compared to a foreign-born population of 12% in 2000, 26% of U.S. based Nobel Prize recipients from 1990-2000 were immigrants. Similarly, immigrants are over-represented among members of the National Academy of Science and the National Academy of Engineering (Levin and Stephan, 1999) and non-U.S. citizens account for 24% of international patent applications from the U.S.

Together, the above facts offer prima facie evidence that a large number of U.S. trained graduates relocate to other countries taking human capital and vast potential with them. This location decision has crucial implications for the U.S. S&E workforce and for the U.S. productivity growth.[5] Finally, in an era where most nations realize the importance of having a well trained S&E workforce, the competition to attract these highly skilled individuals has become more and more fierce. Many countries are now engaging in the intense global competition to attract internationally mobile human capital by redesigning their immigration regimes. The UN World Population Policies Database reports that in 2013, approximately 40% of the 172 UN member states declared an explicit interest to increase the level of high-skilled migration into their

[5]For example, Xu (2015) estimates that the impact of doubling the number of immigrants from every non-OECD country would boost U.S. productivity growth by 0.1 percentage points per year.

countries either by attracting foreign or retaining native talent, whereas this share was only about 22% in 2005. Highly developed countries lead this global trend – two thirds of OECD nations have implemented or are in the process of implementing policies specifically aimed at attracting high-skilled migrants. Moreover, many countries in the Asia-Pacific region have been investing heavily in R&D and now collectively perform a larger share of global R&D than the United States (National Science Board (NSB), 2014). Since S&E skills are portable, these global changes not only have implications for the location choice, but also have implications for cross-border transmission of knowledge and for the U.S. advantages in S&E.

In this chapter we leverage the International Survey of Doctoral Recipients (ISDR) dataset to answer a variety of questions that are relevant for the formulation of policies pertaining to scientific workforce development and high-skill migration. The unique nature of the data alleviates many challenges facing the research community studying high skilled emigration from the U.S. Our analysis identifies a number of factors that are relevant for the location choice. The economic performance of the destination country vis-á-vis United States matters for the location decision. At an individual level, the strength of ties to the U.S. versus the destination country (through legal residency and personal/professional networks) appear to play important roles. We also find that the quality of job-skill match is important for location choice. Significantly, we find that the foreign-born doctoral graduates who leave the U.S. are positively selected on the basis of their talent as measured by the quality of the programs they have attended. In addition, the positive selection is purely driven by the choice of graduates who came to the doctoral program from low/middle income countries and the effect is more pronounced for those who have chosen to emigrate back to low/middle income countries. The effect is also strong for those students who have opted for an academic job. We also find that the magnitude of the positive skill selection has increased in the recent decade suggesting a possible trend where the U.S. may be losing the best of the U.S. university trained foreign-born graduates to other countries in the global race to attract talent.

In the second chapter of this dissertation shifts its focus to the matching behavior of doctoral students in the U.S. high skill labor market. In particular, we study if the type of firm a doctoral student in the U.S. chooses to match with has any implications on how productive they are in the match. In a labor market where workers are heterogeneous in terms of their skill they have and firms differ in terms of the skill they require, productivity at the firm level as well as at the aggregate level may depend upon how the market assigns the workers across jobs. For the specific case of scientists in the U.S., the way the labor market sorts may have potential consequences on the the generation of knowledge and innovation in the economy. However, the first step to gain a deeper understanding of the effects of mismatch on match output and subsequently on aggregate variables requires the researcher to quantify this effect. Empirically doing so is a challenging task. This chapter seeks to meet this challenge by offering a transparent theoretical approach that demands relatively little from the data and yet can offer a reliable measure of job-skill match - both at the sectoral as well as at the aggregate level. The measure that we offer is also suitable for identifying the effects of job-skill match on aggregate outcomes.

A fully estimable model that quantifies the effect of match quality on match output was recently put forward in a seminal contribution by Hagedorn et al. (2017). However, the implementation of their methodology comes at a cost. The estimation process is computationally intensive, it places a lot of demands from the observed data such as information on worker and firm characteristics, and information on workers over a period of time who switch employers and finally relies heavily on wage data which known to be noisy and mis-measured. Crucially, the implicit assumption in HLM is that all mismatch occurs purely due to search frictions. This may not be true in the case of observed labor markets as there may be a variety of factors that influence why one observes an individual with a particular skill type employed in a firm with a different skill type. The aim of this chapter is to provide an alternative approach to quantifying the effect of match quality on output that demands less from the data, is less computationally burdensome, and allows for a broader view of what drives the mismatch.

In order to do so, we use the Choo and Siow (2006) matching model of the marriage market and adapt it to estimate the output or surplus from a match between a worker and a firm while imposing little demand on the data. The estimation in our case simply requires information on the skill types of firms and workers and the observed matching pattern between these two groups of agents. Furthermore, this methodology allows us to quantify the effects of job-skill match while remaining agnostic about its sources. We use this model to study the output gains generated by a job-skill match in the labor market for scientists in the U.S. The data comes from the Survey of Earned Doctorates and the Survey of Doctorate Recipients (SDR), which contains detailed information on the educational history of Ph.D. students graduating from

U.S. institutions and the jobs they hold allowing us to infer skill types of workers and firms. We exploit this information and estimate the temporal patterns in the output gain associated with job-skill match by type/industry.

Our analysis offers a number of insights. We find evidence that a job-skill match results in a larger value created as compared to a skill mismatch for all major skill types/industries. There is a large variation in the job-skill match surplus across industries and this surplus varies over time. It is also the case that the a firms' net benefit from matching with a worker of its own type is correlated with economic conditions and the magnitude of the correlation also varies across time as well as across industries. We find evidence that during periods of high economic activity, certain industries (such as Computer Science) are more open to exploit the benefits from cross-type matches as compared to other industries. This behavior possibly captures an industry's propensity toward diversification and innovation. Since the empirical application is focused on doctorate degree holders in science and engineering fields the interpretation of the output or surplus arising from skill match could potentially be production of knowledge that are key to innovation and sustained growth. With this in mind we explore whether the surplus generated by scientists by staying within their own field has any impact on frequency of innovation at the aggregate level. We find that an increase in the aggregate match surplus in the labor market for scientists increases innovative activities in the economy, as measured by total patent applications. This points to a channel through which better job-skill match in the labor market for scientists may provide tangible benefits to innovative processes in the U.S.

The third chapter of the dissertation revisits the issue of skill selection in the out-migration of foreign-born U.S. PhDs. The analysis conducted in the first chapter suggests that foreign-born doctoral graduates who leave the U.S. are positively selected on the basis of their talent as measured by the quality of the programs they have attended, which is the first evidence in favor of positive skill selection amongst highly skilled emigrants. However, as results from OLS regressions, the estimates are partial correlations and may not represent the true relationship between top program attendance and propensity to emigrate. As such there may be many unobserved individual level factors that affect both top program attendance and propensity to emigrate and bias the estimated coefficients. In this chapter, we try to uncover the causal effect of attending a top program on the probability of leaving by instrumenting top program attendance at the individual level by the average past top program attendance from the students' country of origin.

The instrument is plausibly correlated with top program attendance through two channels. Firstly, the presence of doctoral students from a particular country of origin allows schools to elicit more information about the quality of students from that country. A larger number of students in the program, then, indicates that this information may be inducing schools to accept more students from the country in question and raises the probability of top program attendance at the individual level. Secondly, for any individual looking to apply to doctoral programs in the U.S., a larger presence of doctoral students from the individuals' country of origin in a top program may induce the individual to apply to that program.

Using this instrument, we find that although the OLS regressions in the first chapter identify the patterns of positive skill selection in out-migration of foreign-born doctoral students correctly, it severely underestimates the effects. The analysis in the third chapter verifies that there is indeed a strong causal relationship between attending a top program and leaving the U.S. following graduation, and this effect is entirely driven by students coming from low/middle income countries. These results bolster the narrative in the first chapter and indicate that there may indeed be some evidence to support the claim that the U.S. is losing top talent to global competitors, especially rapidly expanding low/middle income countries.

To summarize, the three chapters of this thesis study the post-graduation career choices of doctoral students in the U.S., a group of individuals who are considered very important for the sustained well-being of the U.S. economy. We explore the factors that determine where these students choose to locate for work following graduation and how their choice of the type of firm they choose to match with impacts their productivity. Wherever possible, we draw implications of these choices on the U.S.' capacity to innovate and continue to maintain it's leadership in the global research and development landscape. This thesis puts forth some novel findings in context of the literature on high-skilled emigration from the U.S. While the analysis conducted in this thesis leaves many questions unanswered, the results that we present should inform and guide policies whose goals are to maintain the quality of the skilled S&E workforce in the U.S. and to ensure that the U.S. maintains an advantage in global scientific research and innovation.

1

TO STAY OR NOT TO STAY: LOCATION CHOICE OF FOREIGN BORN U.S. DOCTORATES

Introduction

The contributions of the Science and Engineering (S&E) or Science, Technology, Engineering and Mathematics (STEM) workforce are essential for creation and adoption of scientific knowledge. According to the U.S. Census Bureau, employment in S&E occupations in the U.S. grew from about 1.1 million in 1960 to about 5.8 million in 2011 at an annualized rate of 3.3%, which is twice the annual rate of growth in total employment for the same period. Foreign-born scientists represent a large fraction of this S&E workforce. According to the American Community Survey (ACS) and the Scientists and Engineers Statistical Data System (SESTAT), 26-27% of respondents employed in S&E occupations during 2010 were foreign born. Foreign-born graduates account for 42-44% of the respondents with a doctoral degree (see Table 1), and 58% of this group have earned their doctoral degrees from an U.S. institution (National Science Board (NSB), 2014). This pool of doctoral graduates has grown rapidly over the last few decades. The share of foreign nationals earning doctorates in Science, Engineering or Health in the United States was about 17% during the decade of 1960. By 2010, this share increased to nearly 40% (See Figure 1), the effect being driven by a large influx of students from low and middle income countries (Grogger and Hanson, 2015).[1] The share of doctorates awarded to foreign-born students in the top tier universities has also grown at a rapid pace (See Figure 2). Foreign-born students now dominate the pool of PhD recipients in many key subject

[1]Source: InfoBrief, National Center for Science and Engineering Statistics, NSF-13-300, October 2012.

areas.[2] It is also the case that the foreign-born doctoral students in U.S. universities are drawn from the top-end of the skill distribution in their home countries and there is compelling evidence to suggest that these individuals make significant contributions to research and innovation as students and as professionals.[3,4]

There is also evidence that each year a large number of U.S. trained highly skilled graduates relocate to other countries taking human capital and vast potential with them. According to 2007 Survey of Earned Doctorates (SED), only 53% of foreign doctorate recipients with temporary visas reported that they have a 'definite plan' to remain in the United States.[5] Since S&E skills are portable, the location decisions of these individuals have crucial implications for the U.S. S&E workforce, cross-border transmission of knowledge, and for the current and the future trajectory of the U.S. comparative advantage in the fields of science and technology. In this paper we use a new set of data - the 2010 and 2013 International Survey of Doctoral Recipients

[2]Bound et al. (2009) report that in 2003, foreign-born U.S. PhDs accounted for: 50% of degrees in Physical Sciences, 67% of degrees in Engineering, 68% of degrees in Economics.

[3]Please refer to Kapur and McHale (2005) for anecdotal evidence supporting this claim.

[4]According to Chellaraj et al. (2008), a 10% increase in the size of foreign graduate students in Science and Engineering (S&E) fields leads to 4.5% increase in the university patent applications and 6.8% increase in the university patent grants. Stuen et al. (2012) reports that having an additional foreign graduate student in S&E departments translates into an average gain of 5 extra articles in the department over the course of a doctoral student's 6-year graduate career. Hunt and Gauthier-Loiselle (2008) suggest that a one percentage point increase in the share of immigrant college graduates increases patent per capita by about 15%. According to Peri (2007), compared to a foreign-born population of 12% in 2000, 26% of U.S. based Nobel Prize recipients from 1990-2000 were immigrants. Similarly, immigrants are over-represented among members of the National Academy of Science and the National Academy of Engineering (Levin and Stephan, 1999) and non-U.S. citizens account for 24% of international patent applications from the United States.

[5]Finn (2010, 2014) offers more direct estimates that are based on Social Security and tax information on foreign doctoral recipients. According to Finn (2010), the two year stay rate of foreign students (with temporary visas) who received their S&E doctorate degree in 2005 is about 67%. In a more recent publication, Finn (2014) reports a slightly higher stay rate for 2006. Still, nearly 28% of graduates have relocated to other countries within two years of graduation and the share rises to 34% within a five year period.

(ISDR) – assembled by the National Science Foundation to analyze the location choice of foreign born doctoral recipients. This data set is unique in a number of respects. In Section 2, we outline the features of the ISDR data and explain its contributions in meeting the challenges facing the current research community in further detail. This data allows us to seek answers to a variety of questions that are relevant for the formulation of policies pertaining to scientific workforce development and high-skill immigration. For example, which individual and country specific factors are important in the foreign-born doctoral graduates' decision to emigrate? Are there are any recent changes in the pattern of emigration among foreign-born doctoral graduates? Finally and most importantly, which segment of the skill distribution amongst this population is the U.S. losing to foreign competition? These are the main set of issues that we address in this paper.

Our analysis identifies a number of factors that are relevant for the location choice of foreign born doctoral students. We find that at the time of the decision, the relative performance of the destination country vis-à-vis United States matters for the location decision – an individual is more likely to stay back in the U.S. if the U.S. enjoys a relatively faster output growth and (in some cases) lower unemployment. These results are broadly consistent with earlier findings, e.g., Grogger and Hanson (2015). In addition, we find that FDI inflows to the destination country and a higher patenting intensity (relative to the U.S.) help to attract talent. At an individual level, the status of legal residency and the status of graduate funding appear to play important roles. For example, students with stronger ties to the U.S. via legal residency (U.S. Citizenship/Permanent Residence) are more likely to stay back. The same holds for students who have received a RA/TA-ship or received a B.A. in the U.S. and the opposite is true for those who have received a fellowship or funds from a foreign source. We also find that the quality of job-skill match is important for location choice.

Our analysis yields a set of results on skill selection in emigration that deserve attention – particularly in the context of an ongoing debate in the existing literature on this issue. For example, by observing attrition from a longitudinal sample of scientists and engineers, Borjas (1989) concluded that the least successful group were the most likely to drop out of the sample, and by inference, leave the United States. Other researchers found support of this result for the general emigrant population in countries such as Germany (Constant and Massey, 2003), Egypt (Gang and Bauer, 1998), and Sweden (Edin et al., 2000). At the same time, there are studies which suggest that those with higher levels of education are more likely to return than those with lower levels (Jasso and Rosenzweig, 1988; Reagan and Olsen, 2000). To reconcile these apparent contradictory findings, Borjas and Bratsberg (1996) reasoned that emigrants can be positively or negatively selected depending on the selection that characterized the original migration flow.

Against these findings, we revisit this selection issue and find that the foreign-born doctoral graduates who leave the U.S. are positively selected on the basis of their talent as measured by the quality of the programs they have attended. This result is robust for the full sample as well as for the sub-samples constructed on the basis of the students' country of origin and their choices of destination. In addition, the positive selection is purely driven by the choice of graduates who came to the doctoral program from low/middle income countries and the effect is more pronounced for those who have chosen to emigrate back to low/middle income countries. The effect is also strong for those students who have opted for an academic job. We also find that the magnitude of the positive skill selection has increased in the recent decade suggesting a possible trend where the U.S. may be losing the best of the U.S. university trained foreign-born graduates to other countries in the global race to attract talent.

The rest of the paper proceeds as follows. In Section 2 we discuss the literature that is closely related to our study and some of the major constraints facing research in this field. Variables used in the analysis and their construction are explained in Section 3. In Section 4, we discuss the results. Section 5 concludes with some remarks.

Related Literature and Challenges

This paper is certainly not the first to study the population of highly skilled migrants in the U.S. The rapidly changing landscape of U.S. higher education has previously drawn the attention of researchers. Some have focused on the determinants of changes over time in the representation of foreign born students among doctorate recipients from U.S. universities (Kapur and McHale, 2005; Bound et al., 2009; Freeman, 2010). Others went on to look at the impact of foreign-born graduates on innovation (Stuen et al., 2012) and on the U.S. labor market conditions (Borjas, 2005; Hunt, 2011). In comparison, little work has been undertaken to understand the factors governing location choice of foreign born U.S. doctoral graduates. Notable exceptions include

Black and Stephan (2007) and Grogger and Hanson (2015). When undertaken, the analysis is based on a set of imprecise information. This is not due to the lack of information on high-skilled emigrants. Rather, the manner in which information were made available rendered little scope for conducting a systematic analysis. To be precise, a systematic analysis of location choice requires simultaneous access to two sets of information on individuals. The first set includes conditioning variables that are based on individual characteristics and the second set must inform about the true location choice of an individual as well as about the characteristics of the host and the destination country around the time of the departure. Heretofore, matching these two sets of information has posed a major challenge to the research community. For example, it is possible to learn a great deal through the Survey of Earned Doctorates (SED) about the characteristics of foreign-born doctoral graduates and about their intentions (regrading whether to leave or stay in the U.S. labor market). In practice, however, the researchers knew little about which individual has left and his/her true location choice. On the flip-side, the seminal studies by Michael G. Finn (Finn, 2010, 2014) offer a scientific method for identifying those foreign born graduates who have left the U.S. workforce. However, this information is not suitable for the analysis because the reported data are in the aggregate form lacking any information on individuals. The reports also do not contain any information on the destination choice of an individual.

The lack of precise information has compelled researchers to make heroic assumptions in the analysis of 'stay versus leave' decisions of the foreign graduates. Using the defense that at the aggregate level the 'intend to stay' responses are good predictor for actual behavior, the existing research (Black and Stephan, 2007; Grogger and Hanson, 2015) has used the temporary visa holders' 'intend to stay' responses in the SED as a proxy for the actual decisions of the foreign born graduates. It is true that the percentage of foreign-born graduates who express their intention to remain in the U.S. (in the SED) tracks the actual one year stay rate in (Finn, 2010) closely. This close correlation between the two aggregate variables does not however guarantee that there exists a close match between the intentions and the actions at an individual level. In fact, in our sample we find that the correlation between 'intend to stay' responses and the 'actual stay' rates is only about 0.67. It is therefore important that we not rely on the stated intentions of the students at the time of graduation. Instead, we should align the characteristics of an individual with his/her true choice. The ISDR data offers such an opportunity by informing us about individuals who have actually left the U.S. workforce.

It is also a customary practice in the literature to assume that the foreign-born graduates are destined to return to their country of birth at the time of graduation. Accordingly the conditioning variables which capture the relative economic environment were constructed around the time of graduation with the birth country as the country of reference. This leaves further room for misaligning incentives with the choice of an individual. Needless to say that in an era of globally integrated labor markets, high skilled workers need not return to their country of birth and the macroeconomic conditions that factor into an individual's decision to move must be those constructed at the time of location and on the basis of the actual location choice.

The newly available International Survey of Doctoral Recipients (ISDR) data alleviate many of the above constraints. The data provide information on foreign born graduates who have left the U.S. workforce along with their current locations. The ISDR data also offer us an opportunity to pin down the time of departure for a large group of individuals in the sample. To this, we add information from other sources to learn more about individual characteristics and also about the characteristics of the destination country. Together, the set of information that is rich enough to render itself suitable for a systematic analysis of location choice.

Data and Variables

For the purposes of our analysis, we make use of the newly available 2010 and 2013 International Survey of Doctorate Recipients (ISDR) data, along with the information contained in the Survey of Earned Doctorates (SED) and the 2010 and 2013 Survey of Doctorate Recipients (SDR). The 2010 ISDR survey was the first to track individuals who settled outside U.S. borders. We merge the data on the respondents from the SDR and ISDR to corresponding data from the SED, which allows us to observe all the demographic variables contained in the SED, in addition to what we observe in the SDR and ISDR. We limit our analysis to foreign born doctoral recipients in the 2010 and 2013 SDR/ISDR for whom we have valid information.

Our measure for emigration from the U.S. is an indicator of whether one is in the ISDR (current job location is outside the U.S.) or one is in the SDR (current job location is within the U.S.). Since every respondent of the ISDR lived outside the United States, we assume they chose to leave the U.S. at some point between graduation and the time at which they were surveyed. Every foreign-born respondent in the SDR currently lives in the U.S. so we assume they have chosen to remain in the U.S. Unfortunately, the ISDR does not offer any direct information about the exact time of emigration. However, we are able to identify the time of emigration using other variables in the data. The ISDR reports the date at which an individual started working on his/her principal job. We compare this date to the date at which he/she received his/her degree. If an individual started working at the job within a two year window from graduation, we set the year of departure equal to the year in which the individual started the job.[6] We use a different strategy for individuals who started on the job more than two years after graduation. The ISDR/SDR data allow us to observe the U.S. legal residence criteria of every individual in the sample. A foreign-born graduate who does not have an H1-B visa, permanent residence or US citizenship, is typically allowed to stay in the U.S. for one year after they graduate. We assume that an individual who does not have any record of a H1-B visa, permanent residence status or US citizenship must have left the US one year after receiving his/her doctorate degree. Put together, we are able to identify the departure date of 88.88% of the sample of those who have emigrated. We exclude the remaining 11.12% from the analysis since we do not know when they have left. This leaves us with a sample of 6,169 foreign-born doctoral recipients from U.S. institutions. Of these, 5,238 were in S&E fields.[7] We treat all emigration as permanent since there is

[6]Using this method, we are able to identify exactly the departure dates of 56% of the individuals who emigrated.

[7]Our sample size is smaller than Grogger and Hanson (2015). This is because unlike Grogger and Hanson (2015), we do not rely on the 'intend to stay' responses in the SED. Instead, our sample

no information in the our data pertaining to circular migration.

The descriptive statistics in Table 2 identify some salient features in the data. For example, in the sample of 6,169 foreign born PhDs, two-thirds (4,113) stayed and one-third (2056) emigrated. Among those who left, the share of individuals who came from high income countries is disproportionately larger.[8] A higher proportion of those who stay in the U.S do not stay in academics and a higher proportion of those who leave go to academic positions. Significantly, among those who emigrated, there is a higher proportion of graduates who "Attended a Top Program". To define a 'top program' we use the classification of highly ranked doctoral programs constructed in Finn (2010) where top programs within broad areas of study (e.g. Physical Sciences, Life Sciences, Mathematics, Economics etc.) are identified using data from the U.S. News and World Reports (USN) ranking of doctoral programs, cross-validated by the 1995 National Research Council rankings of doctoral programs. For each of the nine degree fields, Finn (2010) reports 20-25 top-rated departments.[9] It is fair to assume that on the average the graduates of top programs or top schools have a higher productivity in their chosen field. This is evidenced in our data by significantly higher current earnings among those from top programs who remain in the U.S. compared with those from programs ranked lower. The earnings differential is approximately

is limited to those foreign-born who are tracked in the SED as well as in the 2010 or 2013 SDR and ISDR. The smaller sample size is likely to affect the precision of our estimates but not the parameter estimates themselves.

[8]We classify countries into the categories on the basis of the World Bank Country Classification by Income Level.

[9]As an alternative, we also classify schools into 'top 10' and '11-40' categories solely on the basis of the 1995 National Research Council rankings of doctoral programs. We use the average of nonzero scores across all 41 ranked programs. See `https://www.stat.tamu.edu/~jnewton/nrc_rankings/nrc1.html#TOP60`. We report the descriptive statistics for these measures in Table 2.

15%. This is even more pronounced for those employed in the academic sector, with the differential being 24.2%. For those employed in non-academic jobs the earnings differential is 9.3%.

The remainder of Table 2 includes other control variables with potentials to influence the location choice. These include citizenship, residency status, parental education, sex, marital status, and whether one obtained his/her B.A. in the U.S. The list also includes the nature of financial support, whether it be through research or teaching assistantships (RA/TA, Fellowship or Funds from a foreign source). As one might expect, individuals with closer ties to the U.S. in terms of citizenship or residence status are among those who stay. Similarly, those who stay are more likely to have received a RA/TA, while those who receive a fellowship or other foreign support are more represented among the leavers. On a scale of 1-3, respondents of ISDR/SDR are asked about the quality of the match between their field of study and their current job. The responses are categorized as 1 being the best match and 3 being the poorest match. We recode the responses such that 1 represents "best match", and 0 otherwise (corresponding to responses recorded as 2 or 3 in the original variable). The variable, "Job in Field in which Trained" in Table 2 indicates that those who leave are better matched than those who stay.

The ISDR allows for a better construction of potentially relevant macroeconomic variables by informing us of the country in which an individual currently resides. Having more exactly identified departure dates and emigration location, we are able to create variables that capture economic climate (relative to the U.S. and around the time of departure) of countries to which individuals has chosen to relocate. For the individuals who have not left the U.S., we simply assume that the relevant comparison country for their location choice is their country of birth. We construct these variables based on information from the World Development Indicators (WDI) published by The World Bank. For example, to construct a relative GDP growth variable, we first standardize the per-capita GDP growth of countries. Next, we take the average of this standardized variable for three years preceding the date of departure. The relative GDP growth rate is then defined as the ratio of the averaged standardized U.S. GDP per capita growth rate to the averaged standardized GDP per capita growth rate of the country to which an individual has emigrated. Following the same procedure, we construct the relative unemployment variable. FDI inflows serve as a proxy for the economic openness and it is simply defined as the lagged-three year average of FDI inflows (in 2005 USD) as the percentage of GDP for each country.

We construct a salary premium variable on the basis of the self reported earnings and job type information that are available in the 2010 and 2013 SDR/ISDR surveys. This variable intends to capture the effects of earning potential on the location choice. We classify the job types into 39 categories according to the Job Code for Principal Job (minor group) classifications in the SDR.[10] We divide the foreign salaries by the PPP conversion factor to exchange rate ratio of the corresponding countries, thus giving us the equivalent U.S. salary in terms of purchasing power.[11] Next, we compute

[10]Some information about cross-country wages/salaries by occupation is available in 'The Occupational Wages around the World Database' published by International Labor Organization. The job classification in this database is completely different from those used in the SDR. As a result, this data does not serve our purpose.

[11]This ratio, also called the national price level, tells us how many dollars would be needed in the country in question to buy a bundle of goods that costs one dollar in the U.S. The ratio trivially equals 1 for the U.S.

average salary for each of the 39 job categories in every work location (country). Our strategy here is compare salaries within a profession for two foreign-born graduates - one who is currently working in the U.S. labor market versus the one who has left U.S. for another country. Accordingly, we define salary premium in a job location for a particular job category as the log difference between the average salary of the job category in that location and the average salary of the same job category in the U.S. However, there is one major caveat in this construction. There are a number of locations (countries) which are sparsely represented in the sample and the salary information for all 39 categories are not available for these countries. This limits our ability to use this variable for the full sample. We are however able to use this variable in the analysis pertaining to the most represented countries such as India, China, South Korea and Taiwan.

We factor in the impact of the R&D environment on location choice by constructing a relative average patenting intensity variable. We define patenting intensity as the total patent applications (direct and PCT national phase entries) per capita and the variable represents the average patenting intensity (relative to the U.S.) over three years preceding the date of departure. We also use the relative rule of law variable in the analysis by holding U.S. as the numerator country. Information for these variables are drawn from the WDI Governance Indicators. The Data Appendix provides additional information about the variables and their sources.

Model and Results

We estimate linear probability models of the form:[12,13]

$$\mathbb{P}(\text{leave}_{ict}) = \alpha + \boldsymbol{\beta_1}\mathbf{X}_{ict} + \boldsymbol{\beta_2}\mathbf{Z}_c + \delta_c + \tau_t + \epsilon_{ict}$$

The variable 'leave' is an indicator for whether one emigrated. **X** is a vector of commonly used co-variates listed in Table 2 [see (Black and Stephan, 2007; Grogger and Hanson, 2015)]. To this we add an indicator of having 'Attended a Top Program' to captures the skill selection pattern in emigration. **Z** is a vector of country specific variables measured around the time of an individual's departure from the U.S. labor market. For those who left, these are the relative (to the U.S.) economic and political conditions in the country to which they have emigrated. For those who did not depart, these are relative conditions at the time of graduation in their country of birth. Finally, δ_c and τ_t are country of birth and PhD cohort fixed effects, respectively.

Column (1) of Table 3 includes all variables from the **X** and **Z** vectors plus a set of cohort fixed effects τ_t. In column (3), we limit the sample to students in the S&E programs. The point estimates of the variable "Attended a Top Program" suggest that those who graduate from top programs are 3.64 percentage points more likely to relocate outside of the U.S. Among S&E workers, the estimate jumps to 3.96

[12] We recognize the limitations of the linear probability model in terms of predictions outside of the 0 to 1 range. However, a logit/probit estimation is beyond our scope due to the potential incidental parameters problem arising from a large number of fixed effects. Moreover, the linear probability model lends itself to easily interpretable parameter estimates.

[13] Our analysis is based on richer set of information than previous studies. Still, our data is not rich enough to draw any causal inference.

percentage point, and both these estimates are significantly different from zero at the 0.05 level. This is the first evidence in favor of positive skill selection amongst highly skilled emigrants. This evidence is at odds with some of the earlier conclusions that are either based on a sample of scientists and engineers (Borjas, 1989), or on samples of the general emigrant population (Constant and Massey, 2003; Gang and Bauer, 1998; Edin et al., 2000). Our results however finds support in Jasso and Rosenzweig (1988) and Reagan and Olsen (2000) who suggest that those with higher level of education (skills) are more likely to return than those with lower levels of education. The result is also consistent with the descriptive statistics generated by Finn (2010) who suggests a lower aggregate stay rate among those from top programs.

The effects of the other variables in the **X** vector are consistent with what one might expect and these effects have been noted by previous studies (Black and Stephan, 2007; Grogger and Hanson, 2015). One is less likely to leave if he/she is closely tied to the U.S. (received a B.A. in the U.S., a U.S. permanent resident or citizen) at the time of graduation. The source of support for graduate studies matters as well, with RA and TA recipients are less likely to leave and those receiving foreign support are more likely to leave. This could be due to the fact that graduates with RA or TA-ships face a larger set of opportunities in the U.S. since they work closely with faculty members and have access to a stronger professional network. These students also represent a better pool within their own doctoral programs. A lower stay rate of students with foreign support could be due to the fact that foreign funding often requires students to return to the country of support after graduation.

We add country specific variables to the analysis with more precision by knowing who actually left, when they left, and their true location choice. As in Grogger and Hanson (2015), our results suggest that countries that are growing slower than the U.S. around the time of graduation are less likely to attract graduates. In addition, we find that FDI inflows and relative patenting intensity matter as well. Graduates are more likely to emigrate to a country if a country attracts more foreign investment and graduates are more likely to stay back if the U.S. patenting environment is better than what prevails in the alternative location. The effect of the patent intensity however is much weaker in magnitude than the other two variables. The effects of relative unemployment appear to be large with correct signs. However, the effects are not statistically significant presumably due to its strong association with other country-wide variables, such as GDP growth. Similarly, the relative rule of law appears with correct sign but without significance. This is not surprising since the relative rule of law variable is very persistent and a sizable portion of our sample is made of students who came from high income countries. For these students, it is likely that the comparison is between the quality of the U.S. rule of law and similar rules of law that prevail intese high income countries. This diminishes the importance of the relative the rule of law variable for the full-sample.

It should be noted that we haven't explicitly included changes in the immigration policy as a control. This may raise some alarm and we defend our decision on the following grounds. All our regressions include cohort fixed effects to control for the set of time variant factors which includes any changes in the immigration policy. In addition, our sample consists of individuals who have emigrated immediately after graduation or at most within two years of graduation. The availability of the H1-B visa is unlikely to have an effect on our sample since doctoral graduates in the S&E field can take advantage of the student visa (F-1) practical training for 1-2 years before initiating the labor certification and H1-B process.[14] It is also the case that a large fraction of the doctoral graduates are interested in either academic and/or R&D related jobs for which H1-B visa cap is less stringent.[15]

In columns (2) and (4), we add birth country fixed effects δ_c which increases the R^2 but reduces the sizes of parameter estimates including the size of the top program coefficients. This is not surprising since country of origin fixed effects hold constant any variation that might occur from people being from different countries. If there is clustering of students from certain countries in top programs and those countries are more likely to have students that leave, then these estimates will understate the potential top program effects. For a clearer picture, we stratify the sample by country of origin. In particular, we divide the sample on the basis of whether the country of origin is a developed versus a developing country while keeping in mind that the influx of foreign-born doctoral students in the recent years is driven by those from middle/low income countries whereas the inflow of students from high income countries has remained relatively stable (see Figure 2). The split-sample analysis yields a number of important insights to which we turn next.

The results from the split samples are reported in column (1) and (2) of Table 4.

[14]Foreign-born graduates are allowed to reside in the U.S. under a Optional Practical Training (OPT) provision, which is linked to their F-1 student status.

[15]To be clear, academic and academic institution affiliated jobs are cap-exempt. In addition, the law exempts up to additional 20,000 foreign nationals holding a master's or higher degree from U.S. universities from the cap on H1-B visas. Still, to make sure, we separately run regressions controlling for H1-B regime changes by including a dummy that takes the value 1 if there was a regime change within two years of an individual's graduation year. The results remain mostly unchanged. These results are available upon request.

Interestingly, we find that the positive skill selection that we uncovered in our baseline analysis is driven entirely by those students from low/middle income countries. In fact, students from high income countries who attend a top program are more likely to stay back, even though this effect is not significant. Thus, while there are an increasing number of students from low/middle income countries populating U.S. doctoral programs, it is the most skilled amongst this group of students who choose to leave the U.S. following graduation. It also appears that some conditioning variables play different roles across the two groups of students. For example, the macroeconomic climate appear to matter more for those from high income countries than for those from low/middle income countries. A possible explanation is that students from low/middle income countries may face a much larger set opportunities back at home with a U.S. doctorate degree and hence are more shielded from economic vagaries. It may also be the case that students from low/middle income countries who come to the U.S. for higher education are typically from the higher income strata at home. High income countries are similar to the U.S. in terms of the skill distribution and other characteristics. As a result, a U.S. degree buys relatively less insurance against the economic downturns in these countries and the general economic conditions at home may matter more for a student from a high income country when deciding whether to return home or not.

Emigration has a direct consequence for the cross-border transmission of knowledge. The scope of such transmission varies the across sectors and it is important that we learn which sectors in destination countries have benefited most from the location choice. Column (3) and (4) indicate that the intensity of skill selection is considerably larger for those employed in the academic sector than in the non-academic sector. The effect is in fact absent for those leaving for non-academic jobs. This is not surprising given the changing landscape of the U.S. higher education sector where there is a trend of hiring non-tenure-track faculty in the place of full time tenure track faculty (Ehrenberg, 2012). Given the difficulties in finding tenure track jobs in the U.S., the academic sector outside the U.S. is more attractive to foreign-born doctoral students - particularly for those graduating from top programs with an inclination toward academic/research positions.[16] This behavior has potential consequences for the transmission of knowledge. Unlike the non-academic sector, there is more scope in the academic sector to disseminate knowledge to a broader audience in a non-rival setup. Therefore it is reasonable to expect that a graduate from a top program in the U.S. is likely to generate a greater diffusion of knowledge abroad when working in an academic environment.

The results so far offer strong evidence in favor of positive skill selection among foreign-born doctoral graduates who have emigrated. Moreover, the result is primarily driven by the group of students who came to the U.S. from low/middle income countries. In Table 5 we pay special attention to students from the countries such as India, China, and Taiwan who dominate the low/middle income group. To this group we add South Korea to obtain the top four sending countries in our whole sample. According to the National Science Board Report 2014 (National Science Board (NSB), 2014), these countries also belong to a set of countries with the most

[16]The SDR data contains information about the nature of academic jobs that the individuals are employed in. We find that a greater proportion of those who have left the U.S are tenured or in the tenure track position as compared to those who stay back. Those who leave are also more likely to report "Applied/Basic Research" as their primary job activity.

rapid expansion of R&D expenditures over the last two decades. For these countries we are able to construct and include the salary premium variable. We also include another endogenous variable indicating the quality of the match between the job and the acquired skills. The baseline results are reported in column (1) of the Table 5. In column (2) we include the job-skill match variable and column (3) includes both the match as well as the salary premium variable. In all the three specifications we find we find strong evidence in favor of positive skill selection. It is also the case that the students from these countries who left for another country have experienced a better job-skill match than their counterparts who stayed back in the U.S. workforce. In contrast, the salary premium variable does seem to not matter in the location decision. We are unable to offer an exact explanation for the absence of this effect. However, it is worth noting that the salary premium variable is constructed on the basis of survey responses on salaries which are known to be noisy. We also find that while relative patent intensity loses its significance, relative rule of law matters in the decision of graduates in this selective sample.

Table 6 reports estimates by splitting the sample of students from low/middle income countries in terms of their chosen destinations. In columns (1) and (2), we report the estimates for the group of graduates who have left for another low/middle income countries. Columns (3) and (4) report results for those who have left for a high income country. Some of the covariates lose significance due to the drop in the number of observations. However, the coefficients on the top program indicator remain robust. In fact, we find this coefficient to be much larger for the group who relocated to a low/middle income country as opposed to those who relocated to a high income country.[17]. As in the previous cases, the job-skill match variable assumes importance with a larger coefficient for the group who have returned to the low/middle income countries.

Put together, the data seem to suggest a number of salient facts. First, a large fraction of U.S. trained foreign-born doctorates leave the U.S. S&E workforce after graduation and there is robust evidence of positive skill selection in this group. This effect is driven entirely by those students who come from low/middle income countries and the top talent amongst this group are more likely to return to a set of low/middle income countries experiencing high growth in GDP and R&D over the last couple of decades. It is also the case that there is more evidence of positive skill selection among the graduates who have returned to be employed in the academic sector with more opportunity to disperse knowledge to a wider audience.

We conclude this section by exploring if the positive selection that we have uncovered represents a recent as opposed to an ongoing trend in the high-skill emigration. In Figure 3 we plot the share of students from top programs leaving the U.S. by cohort. The most striking feature of the plot is the sharp upswing in the leave rate since the mid 1990s, which roughly coincides with the slowdown in the U.S. academic labor market and with the rapid growth in R&D and investment in a select few low/middle income countries from which the U.S. universities receive most of the foreign-born students. The lack of information on some of the variables such as unemployment rates, patent intensity, salary premium, job-skill match, and the rule

[17] Our data suggests that six countries (India, China, South Korea, Malaysia, Taiwan and Thailand) out of the total 84 low/middle income countries have attracted 50% of the students who are originally from the low/middle income countries and have attended top doctoral programs in the U.S

of law prohibits us to repeat the original exercise by decade. We separately estimate a reduced baseline specification for the four decades - 1970-80, 1980-90, 1990-2000, and 2000-2011 - without these variables. In Figure 3 we plot the corresponding point estimates on the "Attended a Top Program" variable. While we recognize that these coefficients are not directly comparable to the ones reported in the main analysis, the sharp increase in the size of the coefficient during the most recent decade is too large to be ignored and raises the possibility that the U.S. may be losing a part of its very highly skilled migrant workforce to competition from other countries and that this is a recent occurrence.

Concluding Remarks

Historically, foreign-born graduates from the U.S. universities have made significant contributions to the U.S. S&E workforce. We must not however ignore the changing global landscape. Many countries are now investing heavily in R&D infrastructure and are actively tailoring their immigration policies to attract talent from abroad. The UN World Population Policies Database reports that in 2013, approximately 40% of the 172 UN member states declared an explicit interest to increase the level of high-skilled migration into their countries – either by attracting foreign talent or by retaining native talent. This share was only about 22% in 2005. Developed countries lead this global trend – two thirds of OECD nations have implemented or are in the process of implementing policies that are specifically aimed at attracting high-skilled migrants (Parsons, 2015).[18] In addition, some high growth countries in the Asia-Pacific region, including China, India, South Korea, Taiwan, Malaysia, have been investing heavily in R&D over the last two decades (National Science Board (NSB), 2014). Incidentally, it is the same group of countries who currently dominates the pool of foreign born doctorates and doctoral candidates in the U.S. universities. Concurrently, there is an emerging trend showing low STEM retention rates and a steady decline in the share of U.S. citizens enrolled and awarded advanced degrees in the fields of Science and engineering (See Figure 1). This has been documented by the National Academy of Sciences (National Academy of Sciences and National Academy of Engineering and Institute of Medicine, 2007) and the same sentiment is echoed in a 2012 report by the U.S. Congress Joint Economic Committee (US Congress Joint Economic Committee and others, 2012). There is even some evidence to suggest that professional STEM vacancies take longer to fill now (Rothwell, 2014).

Putting it together, the aforementioned developments seem to have implications for the U.S. S&E workforce. But to be sure, we must dig deeper into the behavior of the foreign-born doctoral graduates who historically have remained a dominant source of supply for the S&E workforce. Such line of inquiry is also important from the perspective of current and future policies. For example, we must learn about the destinations of foreign-born doctorates leaving the U.S. workforce and the direction of the cross-border transmission of knowledge so that appropriate policies

[18]In addition to making these broad changes, many countries have initiated specific programs to promote the return of STEM talent back to their home countries. Examples of such programs include Horizon 2020 (Europe), 1000 Talent Program (China), Brain Return 500 (South Korea), Reverse Brain Drain (Thailand). There is also anecdotal evidence that academic institutions in some countries, such as China, are now paying a very large premium to attract U.S. trained doctoral graduates back to home country.

can be drawn to sustain the U.S. global advantage in science and technology. It is also important that we uncover the recent changes in the pattern of emigration among the foreign-born doctoral graduates, be informed about the individual and country specific factors that are important for the foreign-born doctoral graduates' decision to emigrate, and most importantly, be aware about which segment of the skill distribution among the foreign-born graduates that the U.S. may be losing to foreign competition. These information are essential for tailoring time appropriate immigration policies for high-skill workers, and these are the main set of issues that we address in this paper.

Our analysis points to a number of salient patterns in the data. For example, we find that foreign-born U.S. doctorates who leave the U.S. are positively selected in terms of skill, as measured by the quality of the doctoral program they attended. Moreover, this effect is driven entirely by those students who come from low/middle income countries and there is a higher propensity for this top talent to choose low/middle income countries with fastest growth in R&D as their choice of work location. There is also some tentative evidence to suggest that out-migration from the top portion of the skill distribution of foreign-born U.S. PhDs has intensified during the recent years.

We recognize that our analysis leaves some important questions unanswered. For example, we are unable to address any issues pertaining to circular migration. We are also unable to break the sample further by stay rates. As a result, our focus here has been on the set of individuals who have left the U.S. workforce immediately after graduation. The two year, five year and ten year stay rates of foreign-born U.S. PhDs.

differ significantly, and individuals who move within a few years of graduation and those who move much after presumably represent two very different set of skills with different implications for the S&E workforce and for the cross-border transmission of knowledge. Despite some of these limitations, the results that we present should inform and guide policies whose goals are to maintain the quality of the skilled S&E workforce in the U.S. and to ensure that the U.S. maintains an advantage in global scientific research and innovation.

Figures and Tables

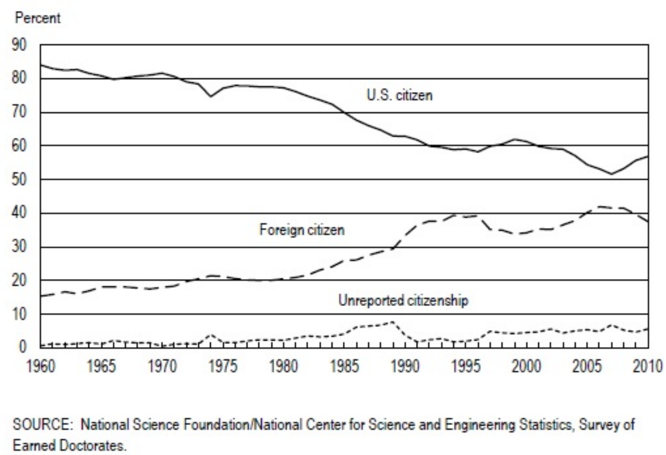

Figure 1.1: U.S. research doctorates awarded in science, engineering, or health, by citizenship: 1960 - 2010

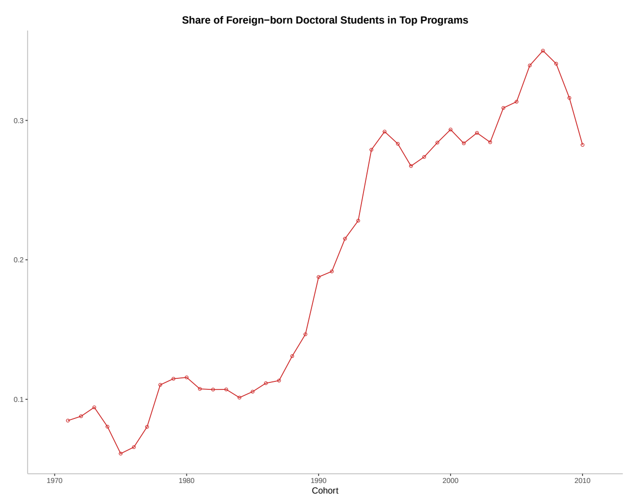

Figure 1.2: Trend in Doctorates awarded by Top Programs

To define a 'top program' we use the classification of highly ranked doctoral programs constructed in Finn (2010) where top programs within broad areas of study (e.g. Physical Sciences, Life Sciences, Mathematics, Economics etc.) are identified.

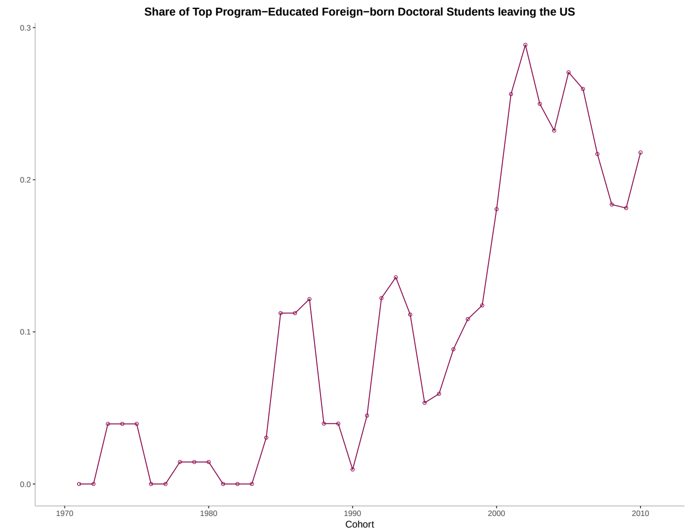

Figure 1.3: Trend in Leave Rates of Top Program Students

Note: Figure 3 is generated on the basis of frequencies weighted by the sampling weights in the ISDR.

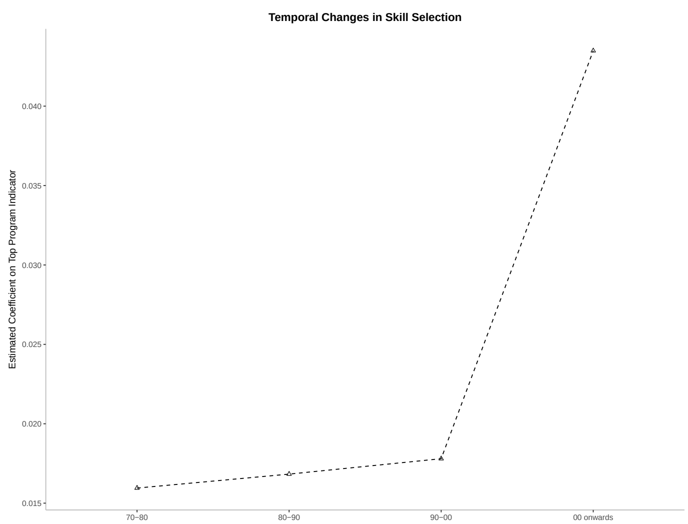

Figure 1.4: Intensity of Positive Skill Selection by Decade 1970-2011

Table 1.1: Foreign-born workers in S&E Occupations, by Education Level: Selected Years, 2000 - 11

	2000	2003		2006		2008		2009	2010		2011
Education	Decennial census	SESTAT	ACS	SESTAT	ACS	SESTAT	ACS	ACS	SESTAT	ACS	ACS
All college educated	22.4	22.6	24.2	23.8	25.3	24.6	24.9	25.2	27.4	26.5	26.2
Bachelors	16.5	16.4	17.7	17.3	18.1	17.2	18.4	18.3	20.1	19.0	19.0
Masters	29.0	29.4	32.0	31.7	33.5	32.7	32.7	33.4	34.9	35.0	34.3
Doctorate	37.6	36.4	37.8	36.6	41.8	37.8	40.9	41.6	41.5	44.2	43.2

ACS = American Community Survey; SESTAT = Scientists and Engineers Statistical Data System.

Sources: National Science Foundation, National Center for Science and Engineering Statistics, SESTAT (2003-10), `http://sestat.nsf.gov`; Census Bureau, 2000 Decennial Census Public Use Microdata Sample (PUMS), and ACS (2003, 2006, 2008, 2009, 2010, 2011).[19]

[19]Notes: This table is reproduced from , Chapter 3, Table 3-27. The data from the ACS and the Decennial Census include all S&E occupations except postsecondary teachers because these occupations are not separately identifiable in the 2000 Census or ACS data files. SESTAT 2006 and 2008 data do not include foreign workers who arrived in the United States after the 2000 Decennial Census and also did not earn an S&E degree in the United States.

Table 1.2: Descriptive Statistics

Main Variables of Interest (Proportions)	Total Sample (n = 6169)	Emigrated (n = 2056)	Stayed (n = 4113)
Currently in High Income Country		75.49	
Currently in Low/Middle Income Country		24.51	
From a S&E Field	84.91	77.58	88.57
From a High Income Country	50.43	64.93	43.17
From a Low/Middle Income Country	49.57	35.07	56.83
Currently in an Academic Job	51.37	61.58	46.27
Currently in a Non-Academic Job	48.63	38.43	53.73
Categorical Covariates (Proportions)			
Bachelors' in the US	16.29	7.73	20.57
Either Parent has a Bachelors'	61.24	59.24	62.24
Male	65.21	69.94	62.85
Married	60.56	53.06	64.31
US Permanent Resident	8.67	1.75	12.13
US Citizen	11.25	2.09	15.83
Received RA/TA	65.08	56.76	69.24
Received Fellowship	21.14	23.39	20.01
Received Foreign Support	5.30	11.04	2.43
Attended a Top Program	33.20	36.43	31.58
Attended a Top 10 School	12.87	14.79	11.91
Attended a School Ranked 11-40	31.14	31.37	31.02
Job in Field in which Trained	71.88	75.78	69.92
Numeric Covariates (Means)			
Age	32.49	32.78	32.34
Relative GDP Growth	-0.14	-0.85	0.20
Relative Unemployment	1.24	1.34	1.19
FDI Inflows to Destination Country	3.57	4.01	3.35
Relative Patenting Intensity	214.75	87.64	278.29
Relative Rule of Law	-0.97	-1.33	-0.79

Table 1.3: Determinants of Leaving the U.S. Following Receipt of a PhD

	All Fields		S&E Only	
	(1)	(2)	(3)	(4)
Measured at time of PhD Receipt				
Attended a Top Program	.03645**	.01416	.03960**	.01612
	(.01653)	(.01320)	(.01600)	(.01255)
Bachelors' in the US	–.04652	–.11983***	–.02975	–.11448***
	(.03572)	(.02325)	(.03812)	(.02343)
Either Parent has Bachelors'	.00759	–.00479	.00769	–.00037
	(.02766)	(.00830)	(.02775)	(.00783)
Male	.03015	.02281	.02560	.01926
	(.02025)	(.01622)	(.01964)	(.01556)
Married	–.09541**	–.04813*	–.10172**	–.05022**
	(.04199)	(.02522)	(.03907)	(.02173)
Age	.00775***	.00325**	.00642***	.00210*
	(.00279)	(.00150)	(.00233)	(.00127)
US Permanent Resident	–.25864***	–.23822***	–.23675***	–.21432***
	(.04840)	(.05111)	(.04719)	(.05001)
US Citizen	–.28957***	–.24660***	–.27255***	–.23371***
	(.03561)	(.03441)	(.03334)	(.03414)
Received RA/TA	–.10482***	–.04342**	–.08379**	–.02571
	(.03137)	(.02001)	(.03457)	(.02197)
Received Fellowship	.00061	.01808	.01608	.02961
	(.02895)	(.02377)	(.03415)	(.02849)
Received Foreign Support	.21735***	.12311***	.25261***	.14519***
	(.05775)	(.03385)	(.06085)	(.03525)
Measured at time of Emigration (US Relative to Destination Country)				
Relative GDP Growth	–.00031***	–.00031***	–.00031***	–.00026***
	(.00009)	(.00010)	(.00008)	(.00008)
Relative Unemployment	.05023	.04939	.05157	.04784
	(.03298)	(.03479)	(.03402)	(.03419)
FDI Inflows to Destination Country	.01014***	.01480***	.01105***	.01550***
	(.00359)	(.00398)	(.00407)	(.00383)
Relative Patenting Intensity	–.00001***	–.00002***	–.00001***	–.00002**
	(.00000)	(.00001)	(.00000)	(.00001)
Relative Rule of Law	–.00012	–.00015	–.00022	–.00024
	(.00040)	(.00025)	(.00041)	(.00026)
Cohort Fixed Effects	Yes	Yes	Yes	Yes
Country Fixed Effects	No	Yes	No	Yes
R^2	.223	.347	.217	.348
No. of Observations	6169	6169	5238	5238

Table 1.4: Determinants of Leaving the U.S. by country of Country of Origin and Job Type

	From High Income	From Low/Middle Income	Academic	Non-Academic
	(1)	(2)	(3)	(4)
Measured at time of PhD Receipt				
Attended a Top Program	–.01542	.03919***	.03679*	–.00467
	(.02038)	(.00658)	(.01961)	(.01314)
Bachelors' in the US	–.13126***	–.07611***	–.16584***	–.05990*
	(.02735)	(.02689)	(.02676)	(.03158)
Either Parent has Bachelors'	–.01323	–.00385	–.00065	–.00461
	(.01358)	(.00722)	(.01620)	(.01686)
Male	.05546**	–.00352	.06572***	–.01607
	(.02663)	(.00771)	(.02282)	(.01551)
Married	–.00894	–.08665***	–.04701*	–.05960**
	(.01937)	(.02746)	(.02556)	(.02539)
Age	.00190	.00406**	.00122	.00630***
	(.00210)	(.00166)	(.00181)	(.00194)
US Permanent Resident	–.33534***	–.14183***	–.25026***	–.20746***
	(.02122)	(.03627)	(.03616)	(.06571)
US Citizen	–.27142***	–.18797***	–.18606***	–.31364***
	(.02820)	(.05237)	(.03052)	(.05439)
Received RA/TA	–.05507**	–.02318	–.03487	–.04334*
	(.02421)	(.02665)	(.02361)	(.02384)
Received Fellowship	.02302	.01176	.00541	.03631
	(.03571)	(.01805)	(.02185)	(.03794)
Received Foreign Support	.12534***	.09177**	.12274***	.11955***
	(.03975)	(.04162)	(.04257)	(.03942)
Measured at time of Emigration (US Relative to Destination Country)				
Relative GDP Growth	–.00038***	.00089	–.00042***	.00090
	(.00009)	(.00073)	(.00012)	(.00072)
Relative Unemployment	.19805**	–.02316	.05875	.04655
	(.08524)	(.02797)	(.04498)	(.03304)
FDI Inflows to Destination Country	.00860**	.02459***	.01536***	.01397***
	(.00422)	(.00493)	(.00464)	(.00394)
Relative Patenting Intensity	–.00048	–.00002***	–.00002***	–.00002***
	(.00029)	(.00001)	(.00001)	(.00001)
Relative Rule of Law	–.00006	–.00023	–.00002	–.00040
	(.00021)	(.00049)	(.00033)	(.00029)
R^2	.318	.372	.330	.404
No. of Observations	3111	3058	3169	3000

Table 1.5: Estimates for Students from India, China, South Korea, Taiwan: Incorporating Match Quality and Salary Premium

	(1)	(2)	(3)
Variables of Interest			
Attended a Top Program	.03495**	.03393**	.03046*
	(.01555)	(.01555)	(.01605)
Job in Field in which Trained		.02627*	.02819*
		(.01417)	(.01483)
Salary Premium			–.02634
			(.01860)
Other Covariates			
Bachelors' in the US	–.08817**	–.08831**	–.10742***
	(.03583)	(.03590)	(.03735)
Either Parent has Bachelors'	–.00921	–.00972	–.01048
	(.01517)	(.01517)	(.01562)
Male	.01684	.01674	.01769
	(.01469)	(.01468)	(.01520)
Married	–.10052***	–.10013***	–.09775***
	(.01690)	(.01690)	(.01755)
Age	.00674***	.00662***	.00628***
	(.00185)	(.00185)	(.00190)
US Permanent Resident	–.11769***	–.11558***	–.11339***
	(.01782)	(.01785)	(.01857)
US Citizen	–.19202***	–.18968***	–.17948***
	(.03413)	(.03415)	(.03564)
Received RA/TA	–.05335*	–.05298*	–.05222*
	(.02883)	(.02876)	(.02998)
Received Fellowship	–.05890*	–.05818*	–.05921*
	(.03221)	(.03217)	(.03358)
Received Foreign Support	.11128*	.11074*	.12977*
	(.06494)	(.06544)	(.06662)
Relative GDP Growth	.00130	.00129	.00022
	(.00150)	(.00152)	(.00119)
Relative Unemployment	–.06738	–.06800	.02953
	(.07607)	(.07606)	(.04647)
FDI Inflows to Destination Country	.02778***	.02766***	.03705***
	(.00387)	(.00387)	(.00474)
Relative Patenting Intensity	–.00082	–.00082	–.00049
	(.00052)	(.00052)	(.00047)
Relative Rule of Law	–.00096**	–.00095**	–.00116**
	(.00044)	(.00044)	(.00047)
R^2	.283	.284	.286
No. of Observations	2867	2867	2672

Table 1.6: Estimates for Students from Low/Middle Income Countries: Where do they go?

	To a Low/Middle Income Country		To a High Income Country	
	(1)	(2)	(3)	(4)
Variables of Interest				
Attended a Top Program	.03222***	.02817***	.01288**	.01487***
	(.00764)	(.00840)	(.00538)	(.00424)
Job in Field in which Trained	.02105***	.02175***	.00755	.01888**
	(.00741)	(.00565)	(.00713)	(.00751)
Salary Premium		–.01042		–.02004
		(.00892)		(.01622)
Other Covariates				
Bachelors' in the US	–.07144***	–.08123***	–.02249	.00377
	(.02218)	(.02364)	(.02273)	(.01796)
Either Parent has Bachelors'	–.01095	–.01399***	.00396	.00864
	(.00770)	(.00463)	(.00595)	(.00658)
Male	.00795	.01546**	–.01659**	–.00964
	(.00584)	(.00601)	(.00683)	(.00574)
Married	–.06127***	–.08503***	–.03852**	–.02558***
	(.02160)	(.01599)	(.01511)	(.00796)
Age	.00344*	.00243	.00083	–.00016
	(.00182)	(.00153)	(.00100)	(.00131)
US Permanent Resident	–.10792***	–.09599***	–.04931***	–.02841***
	(.03527)	(.03281)	(.01627)	(.00870)
US Citizen	–.13415**	–.14114**	–.07757**	–.04945
	(.05445)	(.06363)	(.03251)	(.03287)
Received RA/TA	–.04328	–.04651	.01962	.03395**
	(.02749)	(.03020)	(.01713)	(.01590)
Received Fellowship	–.01429	–.02908	.04024**	.03830***
	(.02336)	(.02199)	(.01854)	(.01407)
Received Foreign Support	.09245*	.06918	.06234	.03054
	(.05054)	(.06194)	(.07666)	(.07559)
Relative GDP Growth	.00011	–.00257*	.00123	.00014
	(.00063)	(.00135)	(.00098)	(.00141)
Relative Unemployment	.08869	.09384	–.31940**	–.26029**
	(.05681)	(.06037)	(.13203)	(.10320)
FDI Inflows to Destination Country	.03450***	.04327**	.02824***	.03251***
	(.01128)	(.02087)	(.00697)	(.01004)
Relative Patenting Intensity	–.00001***	–.00006	–.00002**	–.00056***
	(.00000)	(.00007)	(.00001)	(.00019)
Relative Rule of Law	.00010	.00027	–.00033	–.00043
	(.00029)	(.00067)	(.00038)	(.00057)
R^2	.359	.405	.488	.600
No. of Observations	2813	2312	2582	2053

2

JOB-SKILL MATCH IN THE LABOR MARKET FOR SCIENTISTS AND ITS AGGREGATE IMPLICATIONS

Introduction

The allocation of resources among competing ends is potentially important for aggregate outcomes.[1] The same holds for a labor market where workers are heterogeneous in terms of their skill they have and firms differ in terms of the skill they require. In this case, the productivity at the firm level as well as at the aggregate level may depend upon how the market assigns the workers across jobs. Given this, a deeper understanding of how to quantify the effects of job-skill match and how such effects shape productivity is central to understanding how this impacts the working of an economy. Researchers studying job-skill match and it's effects are however faced with a fundamental question – *"How can the gain (loss) due to skill match (mismatch) between heterogeneous workers and firms be quantified?"* The present paper seeks to address this question by offering a transparent approach that demands relatively little from the data and yet produces reliable estimates of the output gain due to a job-skill match - both at the sectoral as well as at the aggregate level. The measure that we offer is also suitable for identifying the effects of job-skill match on economic outcomes.

Empirically estimating the effect of a job-skill match on output has proven to be a challenging task. The current workhorse model of labor market matching with heterogeneous agents that allows for mismatch is due to Shimer and Smith (2000), who introduce search frictions into the matching model of Becker (1973). Until recently,

[1]For example, there is large evidence to suggest that misallocation of factors can lower factor productivity and can explain persistent cross-country variation in GDP per capita (Banerjee and Duflo, 2005; Restuccia and Rogerson, 2008; Hsieh and Klenow, 2009; Bartelsman et al., 2009; Jeong and Townsend, 2007).

both the Becker (1973) model and its frictional counterpart in Shimer and Smith (2000) were thought to be empirically unidentifiable. In a recent contribution Hagedorn et al. (2017) (henceforth HLM) show how to identify and estimate components of the Shimer and Smith (2000) model including the match production function from observed data. Having recovered the match production function, HLM generate estimates of the output lost due to mismatch by comparing the estimated production function to a benchmark model that has no mismatch, i.e. the frictionless model of Becker (1973).

While HLM offers a breakthrough in estimating matching models of the labor market, the implementation of their proposed methodology comes at a cost. Firstly, The estimation process is computationally intensive. It involves generating a global ranking of workers and firms which uses an approximation of an algorithm that is NP-Hard. Secondly, the estimation process demands a lot from the observed data such as information on worker and firm characteristics, and requires information on workers who switch employers over a period of time.[2] Additionally, the identification relies heavily on wage data which known to be measured with substantial noise. Most importantly, the implicit assumption in models such as HLM is that all mismatch occurs because search is costly and absent these frictions the model would approach the perfect sorting equilibrium of Becker (1973). In reality, there may be many underlying factors starting from remuneration structures across skill types, sector specific skill shortages (or over-supplies), varying growth opportunities across sectors etc. that could shape the way matches are formed. The estimated output lost due

[2]Observing the same worker in different firms is crucial for the ranking algorithm to work.

mismatch in HLM then represents only that part of the output which is lost due to search frictions in the labor market.

The above discussion points to the need for an alternative way of quantifying the output lost due to mismatch that (i) demands less from the data, (ii) is less computationally burdensome, and (iii) allows for a broader view of what drives the mismatch. With this in mind, we lean on the seminal work by Choo and Siow (2006) (henceforth CS) that was originally formulated keeping marriage market matching in mind. We modify their model in such a way that we are able to estimate the output surplus from a job-skill match between a worker and a firm while imposing minimal demands on the data. The estimation in our case simply requires information on the skill types of firms and workers and the observed matching patterns between these two groups of agents. Furthermore, this methodology allows us to quantify the effects of job-skill match on match output while remaining agnostic about its sources. The details of the model are outlined in Section 2 of the paper.

We use this model to study the output gains generated by a job-skill match in the labor market for scientists in the U.S. The data comes from the (licensed version of) Survey of Earned Doctorates and the Survey of Doctorate Recipients (SDR), which contains detailed information on the educational history of Ph.D. students graduating from U.S. institutions and the jobs they hold. Based on this information we are able to construct job-skill matching patterns for high skilled STEM workers over the period 1975-2011. We exploit this information and estimate the temporal patterns in the output gain associated with job-skill match by type/industry. Our analysis offers a number of insights. We find evidence that a job-skill match results in a larger value created as compared to a skill mismatch for all major skill types/industries. There is a large variation in the job-skill match surplus across industries and this surplus varies over time. It is also the case that the a firm's net benefit from matching with a worker of its own type is correlated with economic conditions and the magnitude of the correlation also varies across time as well as across industries. For example, during favorable economic conditions, certain industries (such as Computer Science) are more open to exploit the benefits from cross-type matches as compared to other industries. This behavior possibly captures an industry's propensity toward diversification and innovation. We are also able capture how such propensity has evolved over time and differs across industries.

Our estimation only uses job-skill match data for individuals with a doctorate degree in science and engineering fields. This opens up a broader interpretation of the output or surplus arising from the match. For example, in our case, the output from skill match could very well encompass non-tangible production of knowledge that are key to innovation. With this in mind we move forward and explore whether the surplus generated by scientists by staying within their own field has any impact on frequency of innovation at the aggregate level. For this purpose, we aggregate the information contained in the sectoral time series to construct a diffusion index representing an aggregate measure of surplus from job-skill match. We find that an increase in the aggregate index increases innovative activities in the economy, as measured by total patent applications. This points to a channel through which the effects of job-skill match could show up in the form of higher productivity.

The rest of the paper proceeds as follows – Section 2 outlines the model. Section

3 discusses the data, the empirical methodology and the results. Section 4 concludes with summary and with some comments on possible extensions.

A Model of Worker-Firm Matching

Firms and workers in this setup are heterogeneous in skill requirement and skill ownership. These workers and firms match over a single characteristic – 'skill type'.[3] Consider a firm x with skill requirement i. The firm has two choices, pair with someone with the skill type i or with someone who has a different skill type $-i$. The firm's choice set can be expressed as $j \in \{i, -i\}$. The payoffs from a match with j can be expressed as:

$$V^x_{j,i} = \pi_{j,i} - \tau_{j,i} + \varepsilon^x_{j,i} \tag{2.1}$$

$\pi_{j,i} + \varepsilon_{j,i}$ is the firm's match payoff, consisting of the deterministic match output $\pi_{j,i}$ and the idiosyncratic value from the match $\varepsilon^x_{j,i}$ where $\mathbb{E}[\varepsilon^x_{j,i}] = 0 \quad \forall j$. The idiosyncratic value can be interpreted as an unobserved (to the analyst) preference parameter. A firm of type i, therefore may choose a worker of type $-i$ for one of two reasons: (i) the output from this match, $\pi_{-i,i}$, is high, or (ii) the firm has an idiosyncratic preference for the worker of type $-i$, i.e. $\varepsilon^x_{j,i}$ is high. The transfer/wage paid by the firm to the worker is given by $\tau_{j,i}$.

[3]Note that 'skill type' for firms and workers are constructed symmetrically. The skill types of workers are constructed according to their doctoral degree fields, and job categories are also grouped according to these fields to construct job types. The types are Agricultural Sciences, Biological Sciences, Health Sciences, Engineering, Computer Science and Engineering, Mathematics, Chemistry, Geological and Related Sciences, Physics, Other Physical Sciences, Psychology, Economics and Related Sciences.

$\pi_{j,i}$ can be interpreted in multiple ways depending on the nature of the firm in question. When the firm is a private firm, $\pi_{j,i}$ can be thought of as the output accruing to the firm from this match. Alternatively this value can be thought of as knowledge created by the match. The model allows for a general interpretation of the deterministic value. The firm chooses the type of worker that maximizes payoff:

$$\max_{j \in \{i,-i\}} \quad V_{j,i} \tag{2.2}$$

The problem is similar for the workers. Consider a worker y of skill type i. The worker faces a choice of working in a firm of the same type i or with one that requires a different set of skills $-i$. The worker's choice set is $k \in \{i, -i\}$. The payoffs from each match is:

$$U^y_{i,k} = \gamma + \tau_{i,k} + \rho^y_{i,k} \tag{2.3}$$

The deterministic payoff from this match is $\gamma + \tau_{i,k}$. The γ term is a fixed utility the worker gets from working and $\tau_{i,k}$ is the workers' earnings from the match. The the idiosyncratic payoff is $\rho^y_{i,k}$ where $\mathbb{E}[\rho^y_{i,k}] = 0 \quad \forall j$. The utility maximizing problem for the worker is:

$$\max_{k \in \{i,-i\}} \quad U_{i,k} \tag{2.4}$$

In general, the total value generated by any pair (j, k) with $j \in \{i, -i\}$ and $k \in \{i, -i\}$

$$\phi_{j,k} = \pi_{j,k} + \gamma + \varepsilon_{j,k} + \rho_{j,k} \tag{2.5}$$

It can be shown that a stable (overall) match maximizes total match output (Chiappori and Salanié, 2016).

Equilibrium

In the market, firms demand skills and workers supply them. The demand for different types of skills by a firm of type i can be written as:

$$\mu^d_{j,i} = \mathbb{P}[j|\pi_{i,i}, \pi_{-i,i}, \tau_{i,i}, \tau_{-i,i}] \times p_i \tag{2.6}$$

where p_i is the share of firms of type i and $\mathbb{P}[j|\pi_{i,i}, \pi_{-i,i}, \tau_{i,i}, \tau_{-i,i}]$ is the probability that a firm i chooses a worker $j \in \{i, -i\}$.

The supply of skills of type i across different firms is:

$$\mu^s_{i,k} = \mathbb{P}[k|\gamma, \tau_{i,i}, \tau_{i,-i}] \times q_i \tag{2.7}$$

where q_i is the share of workers of type i and $\mathbb{P}[k|\gamma, \tau_{i,i}, \tau_{i,-i}]$ is the probability that a worker of type i pairs with a firm of type $k \in \{i, -i\}$.

In equilibrium, transfers adjust to clear the market. In general for any pair (j, k) with $j \in \{i, -i\}$ and $k \in \{i, -i\}$, the market clearing conditions are given by $\mu_{j,k} = \mu^d_{j,k} = \mu^s_{j,k}$. Therefore, in equilibrium:

$$\mathbb{P}[j|\pi_{i,k}, \pi_{-i,k}, \tau^{eq}_{i,k}, \tau^{eq}_{-i,k}] \times p_k = \mathbb{P}[k|\gamma, \tau^{eq}_{j,i}, \tau^{eq}_{j,-i}] \times q_j \tag{2.8}$$

We now specialize the model by assuming - (i) $\varepsilon_{j,i}$ and $\rho_{i,k}$, are draws from IID Type I Extreme Value distributions with scale parameter equal to 1, and (ii) the transfer for paid to a worker of a different skill type is zero, i.e $\tau_{-i,i} = \tau_{i,-i} = 0$.[4]

Under these assumptions, choice probabilities can be expressed analytically (MCFADDEN, 1974):

$$\mathbb{P}[j|\pi_{i,i}, \pi_{-i,i}, \tau_{i,i}, \tau_{-i,i}] = \frac{\exp(\pi_{j,i} - \tau^{eq}_{j,i})}{\exp(\pi_{i,i} - \tau^{eq}_{i,i}) + \exp(\pi_{-i,i})} \tag{2.9}$$

$$\mathbb{P}[k|\gamma, \tau^{eq}_{j,i}, \tau^{eq}_{j,-i}] = \frac{\exp(\gamma + \tau^{eq}_{i,k})}{\exp(\gamma + \tau^{eq}_{i,i}) + \exp(\gamma)} \tag{2.10}$$

In equilibrium, expressions (6), (7), (9), (10) form the basis of our identification of the surplus value generated when there is a skill match as compared to a skill mismatch in a worker-firm pairing.

The Surplus Value Generated by a 'Skill-Match'

Let us define the output gained due to a 'skill match' between a worker and a firm for skill type i as:

$$\Pi_i \equiv \pi_{i,i} - \pi_{-i,i} \tag{2.11}$$

$\pi_{i,i} - \pi_{-i,i}$ is the systematic additional output generated by a firm from pairing with a worker with the same skill set i as compared to a worker with a different skill set $-i$. Assuming that the market clears at all times, this quantity can be uniquely identified

[4] This normalization is made to correct for the indeterminacy in identifying the true transfers that clear the market. Note that an across-the-board increase/decrease in transfers will not change the preference orderings of the firms/workers. As a result, any set of transfers that preserve the preference ordering produces the same choice probabilities. Only differences in transfers are identified. Such normalizations are common in discrete choice models. For a detailed exposition see Graham (2011).

from observed data.[5]

A little bit of algebra establishes our identification claim. Note, from the demand condition (6) and the firms' choice probabilities (9), we can express the log-odds of a firm of type i choosing to match with a worker of type i versus type $-i$:

$$\ln\left(\frac{\mu_{i,i}}{\mu_{-i,i}}\right) = \pi_{i,i} - \pi_{-i,i} - \tau_{i,i}^{eq} \tag{2.12}$$

Similarly, using the supply condition (7) and workers' choice probabilities (10) we can write the log-odds of a worker of type i choosing to match with a firm of type i versus type $-i$:

$$\ln\left(\frac{\mu_{i,i}}{\mu_{i,-i}}\right) = \tau_{i,i}^{eq} \tag{2.13}$$

The previous expressions map the theoretically constructed log-odds to observed frequencies of matches. Adding these two expressions and re-arranging gives the main identification result:

$$\Pi_i \equiv \pi_{i,i} - \pi_{-i,i} = 2\ln(\mu_{i,i}) - \ln\mu_{i,-i} - \ln\mu_{-i,i} \tag{2.14}$$

Discussion

Expression (14) provides a simple parameter-free estimator for the systematic output gain due to a skill match as compared to that of a skill mismatch in a worker-firm pairing. To fix ideas, consider the market for skill type 'Physics'. We observe (i)

[5] Note that this is equivalent to identifying the systematic output loss due to a *skill mismatch* between a worker and a firm, which is simply the negative of Π_i.

graduates with skill type 'Physics' pairing with jobs of type 'Physics' [(i,i) pairings], (ii) graduates with skill type 'Physics' pairing with jobs of type 'Not Physics' [$(i,-i)$ pairings], and (iii) graduates with skill type 'Not Physics' pairing with jobs of type 'Physics' [$(-i,i)$ pairings]. The observed frequencies of these pairings uniquely identify Π_i. This expression provides us with a quantifiable value of the output gained due to skill match. The statistic in (14) is unit free and as such is comparable across different skill types i.

The estimator in (14) hs a straightforward interpretation. The left hand side of the expression is the total systematic output surplus generated from a skill match. The right hand side is increasing in the observed frequency of (i,i) type pairings, and decreasing in the frequencies of cross-type ($(i,-i)$ and $(-i,i)$) pairings. By revealed preference, if we observe a large number of (i,i) type pairings relative to cross-type pairings then we can conclude that this is due to higher systematic gains from matching within own type. Similarly, if we observed relatively smaller number of (i,i) type pairings as compared to $(i,-i)$ and $(-i,i)$ pairings, revealed preference leads us to conclude that the output produced from matching with another skill type is relatively higher than matching with the own type. Thus, the observed matching behavior of agents in the model inform us about the surpluses generated from different types of matches. Expression (14) captures this succinctly.

When would a particular firm be more likely to hire a worker of a different skill type? Typically, we expect that firms match with someone of a different skill type when they are trying to diversify their production activities/engage in the production of new products which requires them to seek out new skills in the workforce. For example, a tech company such as Google may require expertise in statistical methodology as a part of their expansion into data analytics. Hence, they may hire Economists or Mathematicians (who own these skills) to meet this requirement. Theoretically, this shows up as an increase in the payoff from matching outside the firms' own type $\pi_{-i,i}$ relative to the payoff from own-type matching $\pi_{i,i}$. We also expect that economic conditions that change firms' incentives to innovate or diversify change the way they demand certain skills in the marketplace which can be captured through the changes in the estimated output surplus from out model. We return to this point in the next section.

The model presented is based on the CS marriage market model and as such, some of the limitations of the CS model applies to ours as well. The other assumption built into the framework is that the market under study is frictionless, i.e. the market clears at all times. While this is a strong assumption in a general labor market setting, in the specific case of the market for doctoral students in the U.S. it is perhaps a reasonable approximation since there is very little unemployment for STEM doctoral students. Similarly, the parametric assumption made in the model in order to derive choice probabilities has been criticized in the literature that followed the CS model.[6] This is because Type I Extreme Value distributions imply the IIA property, i.e. it rules out any correlation in unobserved tastes across different types. In the present case, the choice set of firms is binary and hence the restrictions imposed by the IIA

[6] Graham (2011) and Galichon and Salanié (2015) who study the Choo and Siow (2006) model explain it's caveats in more detail. In perhaps the most significant update to the methodology in the Choo and Siow (2006) model, Galichon and Salanié (2015) show that the general matching model is identified even when the distributional assumptions of Choo and Siow (2006) are completely dispensed with, but estimation is not as straightforward since it requires the estimation of distributional parameters. We choose to use the current framework because of its simplicity.

property are inconsequential. For workers, the choice probabilities depend only on transfers in own type matches due to the normalization performed. However it is easily shown that the probability that a worker chooses a firm of it's own type is increasing in the value of the transfer that the worker receives from an own type match. In a binary choice setting, the probability that the worker chooses a firm of a different type is simply one minus the probability that the worker chooses a firm of its own type. As such, this characterization is sufficient to capture the effect of transfers on different job choice probabilities of workers. Finally, the model is static in nature, and we apply this to compute the equilibrium surplus values from a skill-match by skill type and by cohort. There are no dynamics explicitly included in the model. However, we recognize that the equilibria from cohort to cohort may be linked to each other and choose to incorporate this in a reduced form sense in the empirical analysis that follows, by allowing Π_i to have an autoregressive component. In the next section, we describe our empirical results from taking expression (14) to the data on job choices of STEM doctoral students in the U.S.

Empirical Results

Data and Construction of Skill Types

The primary data sources for this analysis are the licensed version of the Survey of Earned Doctorates (SED) and Survey of Doctorate Recipients (SDR), both conducted by the National Science Foundation. The SED is an annual survey of individuals receiving a research doctorate from an accredited U.S. institution in a given academic year, containing information on the doctoral recipient's educational history and demographic characteristics. The SDR provides, over and above that contained in the SED, demographic, education, and career history information from individuals with a U.S. research doctoral degree. Between the SED and the SDR, we are able to observe research doctorates' fine field of degree (FFOD), and the job category that best describes her principal job at the time the SDR was conducted. Our sample is constructed on the basis of the individuals surveyed in the 1995, 1997, 1999, 2001, 2003, 2006, 2008, 2010 and 2013 SDR.

Central to our analysis is the definition of 'types', i.e. the skill type of the worker and the skill requirement of the job. The skill type of scientists in our data is a direct analogue to their field of degree. The SED/SDR classify doctoral degrees into 352 types. However, given the sample size, estimation of (17) on the basis of all these types is problematic since we would observe a lot of zero cells in the match distribution. Additionally, it isn't clear how much more information is added to the analysis by a finer classification of degree fields, since many of these fields that fall under a broad degree type presumably provide the worker with very similar skills. Hence, we group these degree fields into 12 groups to generate broad skill types. We also restrict our attention to STEM fields alone, dropping Humanities from the analysis. The types are Agricultural Sciences, Biological Sciences, Health Sciences, Engineering, Computer Science and Engineering, Mathematics, Chemistry, Geological and Related Sciences, Physics, Other Physical Sciences, Psychology and Economics and Related Sciences. Appendix A describes this grouping.

Constructing the corresponding job types is less straightforward. The SDR provides a "Job Category" list and asks respondents to choose the category that best describes the job they hold. These job categories, however, aren't grouped by degree field and as such does not provide an easy correspondence with the skill type of the workers that we construct. Furthermore, these job categories change in a non-systematic way between survey years.[7] Hence, we resort to manually classifying job categories in every SDR survey year to the 12 broad groups to produce the corresponding skill types for jobs. Appendix B provides an example of the classification of job categories into groups for the survey year 2010.

On the basis of these classifications, we have a skill type for every individual and the skill type of the job she is matched with. Across surveys, this amounts to 60,187 unique matches. However, there is one important caveat. The SDR collects information about the job held *during the survey date*. For a given individual being surveyed, the current job may not be the first job that they held, especially if they appear in the survey many years after they graduate. Since we are interested in estimating (17) by cohort and the time variation in the added value generated by a skill match is crucial to this paper, we must restrict ourselves to counting only "first matches".[8] That is, by cohort, we only keep those individuals for whom we can say with a degree of certainty that this was their first job. We achieve this by comparing Ph.D. graduation dates and job start dates, and keeping only those individuals for whom the job start date is within two years of the graduation date. This leaves us

[7] Although most of the job categories stay the same, some categories are removed and others are added over different survey years, and these changes need to be kept track of.

[8] For example, suppose we observe an individual in the 2010 survey who graduated in 1995. If we did not filter for a "first match", then this match observed in 2010 would be counted towards a match in 1995, resulting in biased estimates of the additional value generated by a "skill match" for her skill type in 1995.

with 36,830 match observations. We also drop graduation cohorts before 1970 since the number of observations for those years is small enough that it generates many zero cells in the match distribution. Our final matched worker-job data has 35,679 observations across 42 cohorts (1970-2011).

Sectoral Analysis of Skill Match Surplus

Figures 1 and 2 plot the time series of the output (surplus) from job-skill match by industries (sectors). A positive value of the surplus represents a higher benefit from matching within its own type as opposed to matching outside the type.[9] From the plots it is evident that there is a large variation in the surplus across industries and over time. However, in most cases the gain from own type matching is readily apparent. Table 1 reports the time series averages and standard deviations of output gained by job-skill match by sector. Focusing on the cross-sectional variation across industries, we find that the average gain from skill match is much larger in the cases of certain industries such as Psychology and General Engineering whereas industries such as Computer Science, Agricultural Science, and Biological Science record lower averages. We suspect that this is due to the differences in skill specificity across industries. For example, in the case of some industries such as computer science, the skill specificity is much weaker (than say nuclear engineering) due to it's wide applications and it's intersections with other disciplines. Accordingly, there is a smaller difference between the output from matching within it own type and the output from matching outside it's core skill set.

[9]Note that the output lost due to skill mismatch is simply the negative of these numbers.

There is also large temporal variation of the match surplus *within* each industry. This could be due to factors that are specific to the industries, differences in the manner in which industries respond to economic conditions, or both. Within our framework, there is little scope to unveil the role of industry specific factors. However, it is possible to uncover the role of economic conditions. For this purpose, we compute the correlations between output gain Π_i and the lagged two year average of GDP growth in the U.S. Table 2 reports these results. Surprisingly, for many industries the correlations appear to be negative. The negative correlation is more pronounced in the case of Computer Science, Physics, and Other Physical Sciences. These results are open to interpretation. One possible interpretation is that industries often seek to expand during favorable economic conditions. While some industries choose to expand their existing product line, others take advantage of 'good times' and seek to diversify and innovate. Diversification and innovation initiatives often require reaching out to other disciplines with the view that the payoff from matching outside firm's own skill type could exceed the payoff from matching within their own skill type. For such industries, the output gain from matching with its own type will appear to be negatively correlated with economic growth. To anchor this interpretation firmly we compute the correlations between match surplus and economic growth (as above) for the decades of 1980-89, 1990-99 and 2000-11. Table 3 reports these results for a selected group of industries where we see patterns that are consistent with the documented evolution of some industries. For Computer Science, the match surplus is positively correlated with economic growth for the the decade of 1980s. However, the correlation turns negative during the following two decades. The computer science industry was in its nascent stage during the 1980s and naturally leveraged favorable economic conditions to hire more computer scientists in order to obtain a foothold. This translates into a positive correlation between match surplus and economic growth for the decade of 1980s. By the next two decades, the industry had widened its scope to diversify as it gained a central role in every aspect of our daily lives. As the products and skills specific to this industry found wider potential applications, better economic conditions opened up the incentive for the industry to to diversify and innovate by hiring skills outside it's own type set. Presumably this explains the large negative correlation for the two recent decades. We observe a similar pattern in the case of Chemistry. In contrast, industries such as Health Sciences and Engineering seem to exhibit no such patterns and in fact may have become even more selective in their matching preferences.

As noted earlier, the output surplus from job-skill match has a broader interpretation in our model and includes production of intangible knowledge which drives productivity growth. We can verify this by drawing a line from the estimated surplus to a measure of innovative activity such as the volume of patent applications filed in the U.S. One could carry out such an exercise by sector given that information on sector specific patent applications is available. While there exists sector specific patent application data there isn't a one-to-one correspondence between the sectors we define and the sectors in the patent data. We therefore rely on economy wide patent application data and adopt an alternate strategy where we first aggregate the information contained in the 12 estimated surplus series by constructing a diffusion index of output surplus in the labor market for scientists. We then use this index in a

VAR setup to explore how the output surplus interacts with R&D output indicators. The following section offers the details pertaining to the construction of the index.

An Index of Economy Wide Match Surplus

The construction of the diffusion index follows from the vast literature on Dynamic Factor Models (DFM) (Sargent et al., 1977; Stock and Watson, 1991, 2002).[10] The core idea behind dynamic factor models is that the information contained in multiple time series can be pooled into a few series, averaging away idiosyncratic variation in the respective time series. The insight comes from the long standing notion in macroeconometrics that a small set of latent variables, called factors, drive fluctuations in different time series.[11] The premise is that the (smaller set of) factors which constitute the index can replace the original set of time series in forecasting, or as recently explored in Bernanke et al. (2005), in estimating large VAR models.

We follow the methodology in Stock and Watson (1991) to estimate a parametric "single index" model (one dynamic factor), where the index is an unobserved variable capturing all systematic variation in the 12 time series of the output surplus by field. The adoption of this methodology is appealing for the following reasons. Firstly, dimension reduction is necessary given the relatively short length of the time series of our model, given that we wish to eventually estimate a VAR system. Efficiently summarizing the relevant information in each series by constructing an index leads to no loss of interpretability while allowing for cleaner estimation in the VAR stage.[12]

[10] For a recent technical overview of the DFM literature see Stock and Watson (2011).

[11] This dates back to the idea of a "reference cycle" in business cycle analysis in Burns et al. (1946).

[12] This is the basic intuition behind the factor-augmented VAR approach popularized by Bernanke

Secondly, it allows us to combine the disaggregated output surplus series to a series that captures the economy-wide output surplus without imposing any added *a-priori* theoretical structure on the analysis. Finally, using a factor model allows us to maintain agnosticism about what really lies behind the output surplus generated. We simply claim that there is a composite latent component that drives these series in some systematic way. The diffusion index model is specified as follows:

$$\Pi_{i,t} = \alpha_i f_t + e_{i,t} \tag{2.15}$$

where f_t is the index/unobserved factor. The α_i are the 'loadings' for each skill type i. $e_{i,t}$ is an idiosyncratic component that is specific to sector i. These disturbances are AR(1):

$$e_{i,t} = \phi_i e_{i,t-1} + v_{i,t} \quad \text{where } v_{i,t} \sim N(0, \sigma_i^2) \tag{2.16}$$

And finally, the index itself is AR(1):

$$f_t = \Gamma f_{t-1} + \eta_t \quad \text{where } \eta_t \sim N(0, 1) \tag{2.17}$$

The index and idiosyncratic terms in (18) are taken to be uncorrelated at all leads and lags. Finally, the idiosyncratic terms $e_{(i,i),t}$ are uncorrelated across i at leads and lags. The model is estimated using classical Maximum Likelihood and Kalman Filtering after being re-expressed as a state space model.[13] Finally, as can be seen in

et al. (2005).

[13]For a detailed treatment we refer the reader to Kim et al. (1999).

Figures 1 and 2, there are some missing values in the estimated time series of output surplus. Even though the Kalman Filter can deal with missing values, we choose to impute these values using exponential weighting.

Tables 4, 5 and 6 list the parameter estimates of the model described in (18)-(20). We find that the loading values in Table 4 are significant and of the correct signs. The only negative loading is the one on "Other Physical Sciences", the series for which we observed a negative average output surplus over the time series indicating that this degree field behaves differently from the rest of the market for scientists. A potential explanation for this is that this skill grouping is the only one that combines many small and disparate fields of study within the Physical Sciences, which may move the variation in a different direction as compared to the rest of the fields. Engineering, Mathematics, Chemistry and Economics seem to exhibit the highest correlation with the index. Table 5 shows the AR(1) parameters for the model. Almost all the AR(1) parameters are significant, indicating the presence of serial correlation in the idiosyncratic variation in most series, with Biology and Other Physical Sciences being the exception.[14] Table 4 lists the estimated variances for every innovation. Figure 3 plots the Kalman filtered (bold line) and smoothed (dashed line) indices.

Perhaps the most important metric for judging how well the index captures the information contained in each series is the Proportion of Variance Explained (PVE) by the Index. PVE is the ratio of the variance contributed to the series by the index to the total variance of the series. Applying the variance operator to (15) and

[14] We could have specified a more general autoregressive structure for the innovations, but the model is already heavily parameterized given the sample size. We tried using an AR(2) specification but the likelihood convergence failed.

manipulating the expression gives:

$$PVE_i = \frac{\alpha_i^2/(1-\Gamma^2)}{\alpha_i^2/(1-\Gamma^2) + \sigma_i^2/(1-\phi_i^2)} \tag{2.18}$$

Table 7 shows the estimated PVE for every output surplus series. The index does a very good job of capturing the variation in most series. Again, the lowest PVE is seen for Other Physical Sciences. Note that it is not necessary for us to model a single index, the methodology allows for estimation of multiple latent factors. However, given the performance of the single index, we choose to proceed with the parsimonious model. In what follows we use the smoothed index series to explore the dynamic relationship it has with aggregate patent applications.

Economy-Wide Dynamic Relationship between Match Surplus and Patenting Activity

We expect the surplus created by scientists matching with jobs of their own skill type to translate into observable increases in research output. In this section we test this hypothesis, focusing on the most commonly used indicator of research *output* – patenting activity. The data on aggregate patenting applications, and patent issues in the United States comes from the USPTO. Since we are interested in scientists' activities, we consider patenting applications as the indicator for increased scientific activity. Since the index series is stationary, while patenting applications exhibit an increasing trend, we use the HP filtered cycles of the patenting variables in our analysis.

We now estimate a VAR model with patent applications and the output surplus index as the two variables. We choose this parsimonious model keeping in mind the tendency of the VAR setup to eat up degrees of freedom as we increase the number of variables in the analysis. We select the lag length on the basis of a few information criteria. The Hannan-Quinn, Schwarz and Final Prediction Error criteria select a lag length of 1 while the Akaike criteria selects the lag length as 3. We use the smaller lag length given the short length of the time series that we are dealing with. The results of the reduced form VAR(1) model are reported in Table 8. The results indicate that the lagged output index is positively correlated with patenting applications after controlling for patent issuance, and is significant at the 10% level. We also find no evidence of reverse relationship, as is borne out by the zero coefficient on patenting applications on the equation for the output surplus index. We test for Granger causality formally and the test rejects the null hypothesis that the output surplus in the labor market for scientists does not Granger cause Patent Application increases, suggesting a causal link between the two variables that runs from the value generated by scientists by staying in their own field to increased research activity.

Figure 5 plots the impulse response of patenting applications as a result of a shock to the output surplus index. A positive shock to the output surplus index shows up as a increase in patent applications after one year and the effects seem to be fairly persistent, only dying out completely 9 years ahead. Thus, we find evidence to suggest that if scientists stay within their own field and hence are more productive, they are able to generate more knowledge on the aggregate which in our context is captured by an increase in R&D activity such as patenting and that the effect of an increase in the productivity due to a skill match in this market is long lasting.

Conclusion

This paper proposes an empirical methodology that can be used to generate estimates of the output gained (lost) by skill match (mismatch) in a labor market setting. In particular, we formulate a tractable model of labor market matching where workers and firms match on skill types based on the Choo and Siow (2006) model of the marriage market. We show how identification of the surplus value generated by a skill match as compared to a mismatch can be achieved using data on aggregate matching patterns, given well defined skill types of workers and firms. The model in this paper also provides for a more general take on what drives matches across types. In this setup, dissmiliar workers and firms may match optimally simply because the systematic payoff from these matches are high, or because there is some unobserved heterogeneity specific to a worker-firm pair that drives these matches. The key identification result is intuitive and provides for an easy and direct way to quantify the value lost due to a mismatch.

We then take our model to data on the job choices of young scientists in the U.S. in order to quantify the surplus value by skill types of scientists (their major degree fields). The model allows us to generate estimates of the surplus value due to a skill match at a disaggregated level – by skill type and over time. Our results suggest that there is indeed an added value generated by a skill match in most major degree fields, and even though the surplus values demonstrate significant time series variation, they are mostly positive. It is also the case that the a firm's net benefit from matching with a worker of its own type is correlated with economic conditions and the magnitude of the correlation also varies across time as well as across industries. For example, during favorable economic conditions, certain industries (such as Computer Science) are more open to exploit the benefits from cross-type matches as compared to other industries. This behavior possibly captures an industry's propensity toward diversification and innovation. We are also able capture how such propensity has evolved over time and differs across industries. We then exploit the time series variation in these series to test the hypothesis that when scientists stay in their own field and presumably generate more knowledge, this corresponds to observable increases in research productivity in the economy. We study the dynamic relationship between the surplus value and an indicator of research productivity at an aggregated level, namely patent applications. Our findings suggest that increases in surplus value due to a skill match predicts increases in patent applications. This points to a channel through which the effects of job-skill match could show up in the form of higher productivity.

The methodology explored in this paper is general enough that it can be applied to any dataset which allows the construction of well defined skill types. It doesn't rely on data on vacancies at the firm level, neither wage data and characteristics of workers beyond what is required to determine their skill ownership. From the specific case in which we apply the model, we find no evidence of a structural break in surplus value, suggesting the market for scientists has remained relatively stable over the years. The findings seem reasonable within context, labor markets for the very highly skilled may behave differently than labor markets for individuals at the lower end of the skill distribution. In particular, we expect technological changes in the economy to induce changes in the the way skills are used across sectors which may be more visible for a labor market which is not as specialized as that for scientists. In future research we intend to conduct a similar analysis using the National Survey of College Graduates (NSCG) conducted by the NSF, which also contains data on Bachelors and Masters students where we expect much larger volatility in the surplus value generated by a skill match. This analysis when applied to this population, opens up the possibility of exploring the effects of major technological changes in the U.S. economy that have redefined the usage of skills, such as the computing revolution or the rise of quantitative finance.

Figures and Tables

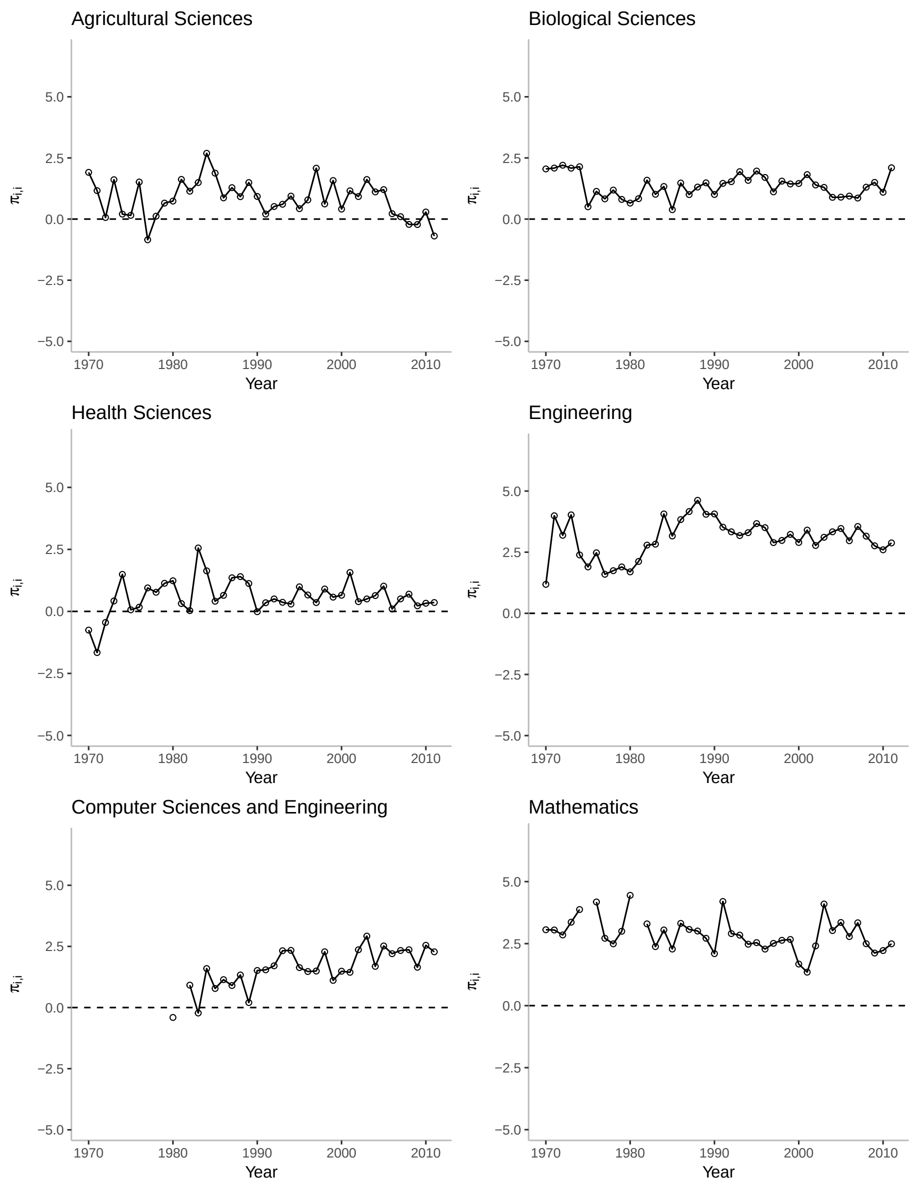

Figure 2.1: Added Value Generated by a Skill Match by Subject Area over Time

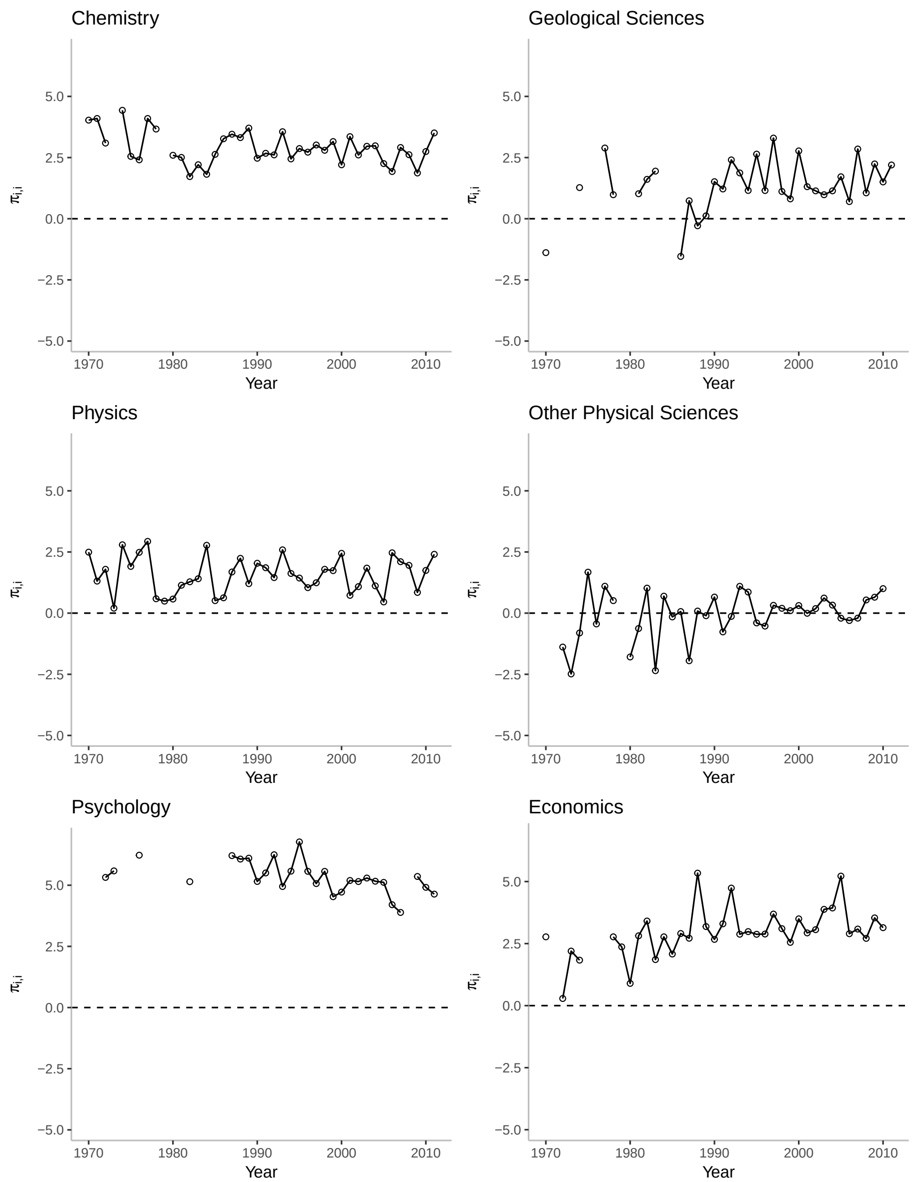

Figure 2.2: Added Value Generated by a Skill Match by Subject Area over Time (Contd.)

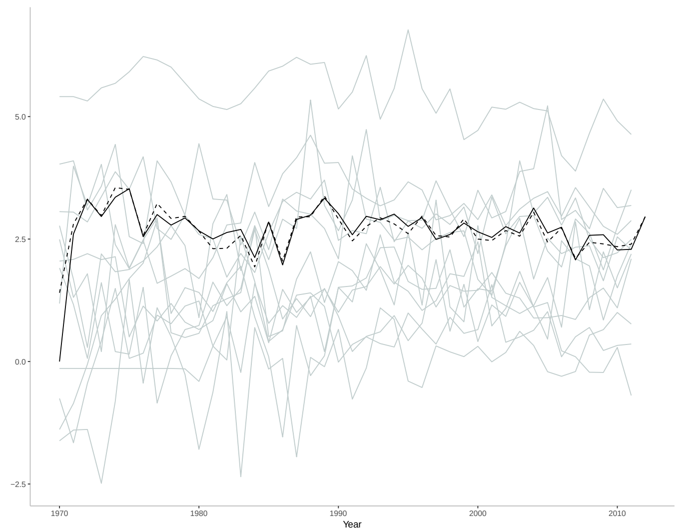

Figure 2.3: Market Index of Added Value Generated by a Skill Match

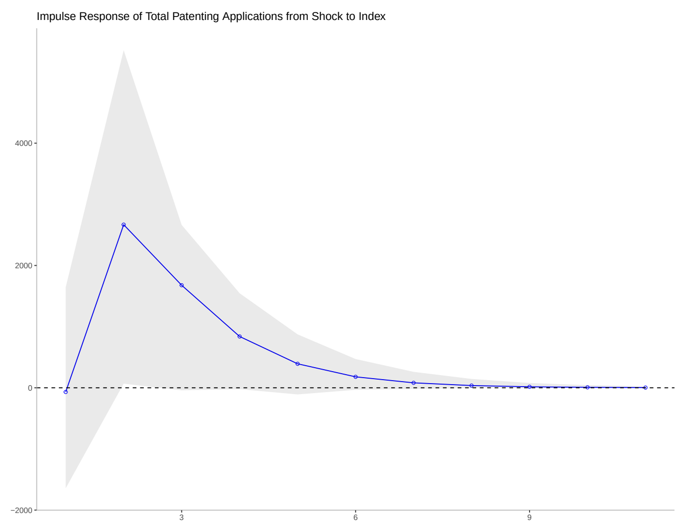

Figure 2.4: Impulse Responses from VAR Analysis

Table 2.1: Added Value Generated by a Skill Match: Summary Statistics

Skill Type	Time Series Mean	Time Series Std. Dev.
Agricultural Sciences	0.8384	0.7535
Biological Sciences	1.3548	0.4729
Health Sciences	0.5912	0.6915
Engineering	3.0539	0.7789
Computer Science	1.5950	0.7976
Mathematics	2.8669	0.6641
Chemistry	2.8962	0.6575
Geological Sciences	1.3376	1.0816
Physics	1.5816	0.7350
Other Physical Sciences	-0.0695	0.9531
Psychology	5.3292	0.6479
Economics	2.9668	0.9639

Table 2.2: Correlations of Surplus Value with Lagged Two Year Average of GDP Growth

Skill Type	Correlation
Agricultural Sciences	0.128
Biological Sciences	-0.077
Health Sciences	0.184
Engineering	0.050
Computer Science	-0.235
Mathematics	-0.041
Chemistry	0.307
Geological Sciences	-0.307
Physics	-0.294
Other Physical Sciences	-0.204
Psychology	0.173
Economics	-0.097

Table 2.3: Correlations with Lagged Two Year Average of GDP Growth: Certain Fields over Time

Skill Type	1980 - 1989	1990 - 1999	2000 - 2011
Health Sciences	-0.322	0.160	0.510
Engineering	0.210	0.098	0.563
Computer Science	0.106	-0.305	-0.437
Chemistry	0.642	-0.113	-0.209

Table 2.4: Parameter Estimates from Dynamic Factor Analysis: Factor Loadings

Parameter	Point Estimate	Std. Err.	z	95% CI	
α_{agri}	0.3683	0.0027	21.1576	0.3630	0.3737
$\alpha_{biology}$	0.5247	0.0023	179.7146	0.5202	0.5293
α_{health}	0.1990	0.0072	11.2894	0.1849	0.2131
α_{engg}	0.8179	0.0021	25.2903	0.8138	0.8220
$\alpha_{computer}$	0.5844	0.0045	11.4559	0.5756	0.5931
α_{math}	0.9599	0.0043	21.6427	0.9514	0.9684
$\alpha_{chemisty}$	1.0775	0.0026	29.8528	1.0723	1.0826
$\alpha_{geology}$	0.4071	0.0029	7.3194	0.4014	0.4127
$\alpha_{physics}$	0.5612	0.0042	87.5493	0.5531	0.5694
$\alpha_{otherphysical}$	-0.0443	0.0078	2.8541	-0.0596	-0.0291
$\alpha_{psychology}$	0.4169	0.0031	104.0651	0.4108	0.4231
$\alpha_{economics}$	0.7961	0.0027	103.9943	0.7907	0.8015

Table 2.5: Parameter Estimates from Dynamic Factor Analysis: AR(1) Parameters

Parameter	Point Estimate	Std. Err.	z	95% CI	
Γ	0.9692	0.0021	12.2791	0.9652	0.9733
ϕ_{agri}	0.3318	0.0027	5.9536	0.3264	0.3372
$\phi_{biology}$	0.6529	0.0053	1.4963	0.6426	0.6633
ϕ_{health}	0.5458	0.0023	16.7598	0.5413	0.5504
ϕ_{engg}	0.8367	0.0019	9.1590	0.8329	0.8405
$\phi_{computer}$	0.8455	0.0048	32.3469	0.8360	0.8550
ϕ_{math}	0.3542	0.0036	2.9772	0.3472	0.3612
$\phi_{chemisty}$	-0.2828	0.0013	81.2008	-0.2853	-0.2803
$\phi_{geology}$	0.6484	0.0016	2.6366	0.6453	0.6515
$\phi_{physics}$	-0.0998	0.0027	4.3641	-0.1050	-0.0946
$\phi_{otherphysical}$	0.0553	0.0025	0.6936	0.0504	0.0602
$\phi_{psychology}$	0.9880	0.0025	9.8228	0.9831	0.9929
$\phi_{economics}$	0.7244	0.0015	102.4458	0.7215	0.7274

Table 2.6: Parameter Estimates from Dynamic Factor Analysis: Innovation Variances

Parameter	Point Estimate	Std. Err.	z	95% CI	
σ^2_{agri}	0.3659	0.0034	5.1609	0.3592	0.3726
$\sigma^2_{biology}$	0.1197	0.0049	3.0320	0.1100	0.1294
σ^2_{health}	0.4196	0.0011	22.2836	0.4174	0.4217
σ^2_{engg}	0.5981	0.0058	3.4286	0.5868	0.6094
$\sigma^2_{computer}$	0.3487	0.0045	9.9409	0.3399	0.3575
σ^2_{math}	0.5047	0.0027	10.4557	0.4995	0.5100
$\sigma^2_{chemisty}$	0.4232	0.0006	10.8221	0.4221	0.4244
$\sigma^2_{geology}$	0.9499	0.0037	15.7425	0.9427	0.9572
$\sigma^2_{physics}$	0.8249	0.0061	9.7988	0.8130	0.8369
$\sigma^2_{otherphysical}$	0.9748	0.0025	21.3985	0.9700	0.9796
$\sigma^2_{psychology}$	0.2707	0.0023	2.6915	0.2663	0.2752
$\sigma^2_{economics}$	0.9303	0.0026	4.4694	0.9252	0.9355

Table 2.7: Proprtion of Variance Explained (PVE) by Common Factor

Skill Type	PVE
Agricultural Sciences	0.8449
Biological Sciences	0.9561
Health Sciences	0.5224
Engineering	0.8470
Computer Science	0.8217
Mathematics	0.9634
Chemistry	0.9766
Geological Sciences	0.6253
Physics	0.8619
Other Physical Sciences	0.0321
Psychology	0.2017
Economics	0.8423

Table 2.8: VAR Results

	Estimate	Std. Error	Pr(>|t|)
Equation: Total Patent Applications$_t$			
Output Surplus Index$_{t-1}$	7301.0881*	3884.0786	0.0680
Total Patent Applications$_{t-1}$	0.3395***	0.1455	0.0092
Total Patents Issued$_{t-1}$	0.0475	0.0306	0.1296
	Estimate	Std. Error	Pr(>|t|)
Equation: Output Surplus Index$_t$			
Output Surplus Index$_{t-1}$	0.227131	0.164479	0.176
Total Patent Applications$_{t-1}$	0.000001	0.000006	0.804
Total Patents Issued$_{t-1}$	-0.000001	0.000001	0.322

3

SKILL SELECTION IN OUT-MIGRATION OF FOREIGN BORN U.S. DOCTORATES: A CAUSAL APPROACH

Introduction

Historically, foreign-born graduates from the U.S. universities have made significant contributions to the U.S. S&E workforce and have been the dominant source of supply of S&E skills in the marketplace. However, against the backdrop of a rapidly changing global landscape of international competition to gain an edge in R&D by attracting top talent, we must gain a deeper understanding of the behaviors of these individuals and the consequences of their choices for the U.S. S&E workforce. The first chapter of this thesis seeks to learn about the destinations of foreign-born doctorates leaving the U.S. workforce and the direction of the cross-border transmission of knowledge between countries. It explores which individual and country specific factors play an important role for the foreign-born doctoral graduates' decision to emigrate and whether there are any recent changes in the patterns of emigration. Most importantly, this chapter seeks to uncover which segment of the skill distribution among the foreign-born graduates that the U.S. may be losing to foreign competition.

The analysis conducted in the first chapter confirms a few existing findings while pointing to a number of salient patterns in the data. The most significant finding of the analysis is that foreign-born U.S. doctorates who leave the U.S. are positively selected in terms of skill, as measured by the quality of the doctoral program they attended. Moreover, this effect is driven entirely by those students who come from low/middle income countries and there is a higher propensity for this top talent to choose low/middle income countries with fastest growth in R&D as their choice of work location. We also find some tentative evidence to suggest that out-migration from the top portion of the skill distribution of foreign-born U.S. PhDs has intensified during the recent years. This hints towards a possible trend where the U.S. may be losing the best of the its university trained foreign-born graduates to other countries in the global race to attract talent.

While the finding of positive skill selection in the out-migration of foreign-born doctoral students from the U.S. is novel in the context of the literature on high skilled emigration from the U.S., we are not able to establish any form of causality.[1] As such, there are reasons to believe that the estimated magnitude of skill selection is biased in simple OLS regressions since there may be many unobserved factors at the individual level which are correlated with both top program attendance and propensity to leave.[2] The direction of the bias is hard to pin down where there are many omitted factors that are correlated with the explanatory variables of interest, and hence there is a possibility that the estimated coefficients understate the true effect of attending a top program on the probability of emigrating following graduation. The goal of this chapter is to resolve this issue. In particular, we seek to empirically investigate the causal link between skill and propensity to emigrate for the sample of highly skilled foreign-born individuals in the U.S. In order to estimate the causal effect of

[1]To reconcile the existing debate regarding positive/negative skill selection in high skilled emigration, Borjas and Bratsberg (1996) reasoned that emigrants can be positively/negatively selected depending on the selection that characterized the original migration flow. However, the argument put forth in Borjas and Bratsberg (1996) requires that the selection in the original migration flow be negative to induce positive skill selection in emigration. Our finding is novel in the sense that we find positive skill selection in emigration even when there is evidence to suggest that the individuals who migrate to the U.S. for higher education belong to the top portion of the skill distribution in their respective countries.

[2]For example, one such factor is parents' socioeconomic status, which may induce individuals to migrate to the U.S. to pursue education and also return home following graduation. While we can proxy for this using parental education, such proxies are imperfect.

top program attendance on probability of emigration involves finding a variable that induces exogenous variation in top program attendance. Absent the availability of any obvious natural experiments that induce such variation, we propose to instrument top program attendance with the *past three year (from year of graduation) average of top program attendance from a student's country of origin* (henceforth referred to as average past top program attendance).

Ex ante, top program attendance at the individual level is plausibly a function of average past top program attendance due to two reasons. Firstly, the presence of doctoral students from a particular country of origin allows schools to elicit more information about the quality of students from that country. A larger number of students in the program, then, indicates that this information may be inducing schools to accept more students from the country in question and raises the probability of top program attendance at the individual level. For example, a program that has had good experiences with Indian graduate students in terms of academic performance and professional achievement may be induced to admit more graduate applicants from India in future cohorts. Secondly, and perhaps less importantly, for any individual looking to choose doctoral programs in the U.S., a larger presence of doctoral students from the individuals' country of origin in a top program may induce the individual to apply to that program. This corresponds to the country-of-origin network effect that increases the chance of application and hence may increase the chances of acceptance into a top program.

Using average past top program attendance as an instrument for top program attendance, we find the first stage relationship between these variables to be as expected – higher the average past top program attendance, greater is the probability that an individual is enrolled in a top program. In comparing the instrumented coefficients with naive OLS coefficients we find that although the OLS regressions in the first chapter identify the key patterns of positive skill selection in the out-migration of foreign-born doctoral students correctly, they severely underestimate the magnitude of these effects. Therefore, the analysis in the third chapter verifies that there is indeed a strong causal relationship between attending a top program and leaving the U.S. following graduation, and this effect is entirely driven by students coming from low/middle income countries. These results bolster the narrative in the first chapter and indicate that there may indeed be some evidence to support the claim that the U.S. is losing top talent to global competitors, especially rapidly expanding low/middle income countries.

The rest of the chapter is organized as follows. The next section describes the construction of the instrument in detail and highlights certain empirical challenges faced in doing so. The following section presents the results and discusses possible reasons for the differences between the OLS and IV estimates. The final section concludes.

Constructing the Instrument

The data for the analysis is the same as in the first chapter – we make use of the 2010 and 2013 International Survey of Doctorate Recipients (ISDR) data, along with the information contained in the Survey of Earned Doctorates (SED) and the 2010 and 2013 Survey of Doctorate Recipients (SDR). When all the information is put together this is a cross sectional, individual level dataset with limited information about the education and work histories of the individuals. For the construction of the instrument, we leverage information on the year of graduation of an individual. To construct the instrument for a particular individual, we simply count the number of top graduates belonging to the individuals' country of origin in the past three years and average them. For example, for an Indian student graduating in 2007, the instrument would be the average number of Indian students who graduated from top programs in the period 2004-2006.

There are certain caveats in the measurement of the instrument. In particular, the data at hand is a (representative) sample of the universe of foreign-born doctoral students in the U.S. and not a census. Therefore, there is a possibility that the data collection mechanism simply doest not sample individuals for certain combinations of graduation year and country of origin owing to the fact that they have very low proportional representation in the population. To fix ideas, consider the example of observing a student from Nepal in 1999. On counting the number of students from Nepal who went to top programs in the years 1996-1998, we find that the counts are mostly zero. There are two possibilities. One – that there were indeed no students from Nepal who attended top programs in that period, or two – that top program attendees from Nepal in that period were not sampled. If the latter is true, the construction of the instrument would be imputing zeroes where the true value of the instrument should be positive. There is also no way for us to tell which of the possibilities arise in practice on a case by case basis.

Under the assumption that the second possibility is true in many cases, the nature of the measurement error in the instrument reduces the variation in the instrument itself. While this may be of concern regarding the strength of the instrument in the first stage, we take solace in the fact that the bias induced by the measurement error in the first stage relationship between top program attendance and average past top program attendance will most certainly be downwards. In other words, the first stage effects that we are likely to find will underrepresent the strength of the true first stage relationship. The other potential concern is that the measurement error is correlated with the country of origin of the student and hence correlates the instrument with the error term in absence of country of origin controls. In all our regressions, we include year of graduation and country of origin fixed effects to control for this correlation.

The problem of zero imputation in constructing the instrument also limits our ability to finesse the measurement of the same any further. For example, we would ideally like to instrument top program attendance of an Indian physics student by the average number of Indian *physics* students who graduated from top programs in the past three years. However, adding this extra layer (field of study) increases the number of possible combinations that need to be counted by an order of 200, while the number of observed top program attendees to populate the counts of these combinations remain the same. This would mean that incorrect zero imputation would become highly likely, severely damaging the instruments ability to induce any variation in top program attendance. We choose to proceed with the cruder version of the instrument in order to keep the potential measurement error to a minimum.

Model and Results

We estimate the following reduced form IV model:

$$\begin{aligned}\mathbb{P}(\text{Attended a Top Program}_{ict}) &= \alpha + \gamma_1 \text{Avg. Past Top Program Attendance} \\ &\quad + \boldsymbol{\gamma_2} \mathbf{X}_{ict} + \boldsymbol{\gamma_3} \mathbf{Z}_c + \delta_c + \tau_t + \epsilon_{ict} \qquad (3.1) \\ \mathbb{P}(\text{leave}_{ict}) &= \delta + \beta_1 1(\text{Attended a Top Program}) \\ &\quad + \boldsymbol{\beta_2} \mathbf{X}_{ict} + \boldsymbol{\beta_3} \mathbf{Z}_c + \delta_c + \tau_t + \epsilon_{ict} \qquad (3.2)\end{aligned}$$

where Avg. Past Top Program Attendance instruments 1(Attended a Top Program). $\mathbf{X}_{ict}$ collects individual specific exogenous variables, $\mathbf{Z}_c$ is a vector of relative (to the U.S.) country specific exogenous variables measured around the time of an individual's departure from the U.S. labor market. Finally, δ_c and τ_t are country of origin and graduation year fixed effects, respectively. Throughout the analysis, we use LIML estimators to alleviate concerns of weak instrumentation.[3]

Table 3.1 reports the first stage regression results for the full sample and the sample that contains only S&E students. As intuition suggests, we find a positive relationship between Avg. Past Top Program Attendance and the probability of Attending a Top Program and the relationship is significantly different from zero. The size of the coefficient is quite small, possibly due to the attenuation caused by measurement error in the construction of the instrument. The bottom of the table reports the first stage robust F statistics. In both samples the F statistic exceeds the rule-of-thumb value of 10. Given that we are using LIML in estimation, this suggests

[3] 2SLS results are very similar to LIML.

that the instrument is strong (Staiger et al., 1997; Stock and Yogo, 2005).[4]

Table 3.2 reports the second stage IV regression results along with the corresponding OLS regressions from Chapter 1 for the full sample and S&E only sample. It is immediately clear that naive OLS underestimates the true effect of attending a top program on the propensity to leave. The instrumented coefficient on the top program indicator is along the same direction as the OLS coefficient, but is about 50-70 times larger than the OLS coefficient. The IV results suggest that those who attend top programs are approximately 60-72% more likely to leave as following graduation as compared to those who don't. In the broader context of skill selection in high skilled emigration, this evidence is a strong contradiction of the earlier conclusions that are either based on a sample of scientists and engineers (Borjas, 1989), or on samples of the general emigrant population (Constant and Massey, 2003; Gang and Bauer, 1998; Edin et al., 2000). The results are consistent with the descriptive statistics generated by Finn (2010) who suggests a lower aggregate stay rate among those from top programs. The rest of the measured effects on exogenous variables remain broadly the same, with some being measured with more precision in the IV regression. Note that both IV and OLS estimates are measured with substantial noise, a pattern that is also noticed in the first chapter. Next, we divide the sample on the basis of whether the country of origin is a high income versus a low middle/income country.

Table 3.3 reports the second stage regression results for the split sample.[5] The

[4]Stock and Yogo (2005) note that the Staiger et al. (1997) rule of thumb may be too conservative when LIML is used. Under the assumption that the F statistic is non-robust, we can compare the estimated values to tables in Stock and Yogo (2005). The non-robust F statistic is 10.04 which exceeds the critical value at size 0.15 (8.96).

[5]The bottom panel of the table reports the first stage partial correlation between the instrument and the endogenous variable. In all regressions, the first stage correlation is strong.

results follow a similar pattern to what was seen in Table 3.2. The instrumented coefficients are in the same direction as the OLS coefficients but are much larger, again suggesting that OLS underestimates the true magnitude of skill selection in the out-migration of the foreign-born. In some cases, the IV coefficients are also measured with lesser noise in the case of the sample of students from high income countries. The IV results suggest that these two groups demonstrate very different behaviors – students from high income countries who graduate from top programs are 53% more likely to stay back in the U.S. following graduation compared to students in other programs while top students from low/middle income countries are almost 1.5 times more likely to leave. The estimated coefficients are also significantly different from zero. The phenomenon of positive skill selection in the full sample is entirely driven by the behavior of students from low/middle income countries. In Table 3.4, we report split sample analyses for S&E students only. The results are broadly the same.

The causal analysis performed thus far serves to bolster the findings from and hence the narrative of the first chapter. Interestingly, we find that although naive OLS regressions identify the patterns of skill selection in the out-migration of foreign-born doctoral students accurately, they underestimate he magnitude of the effect significantly. The estimated IV coefficients are at least 35 times and at most 45 times larger than OLS coefficients. We now briefly consider why this may be the case.

The first possibility is that the OLS results are indeed biased downward. However, there may be a myriad of unobservable individual characteristics that may be correlated with both top program attendance and propensity to leave. As such, finding an economic narrative that supports the downward bias of OLS regressions amounts to discerning which factors create the downward bias, which is next to impossible and foolish to try. We note that the OLS estimate is the partial association between probability of leaving and top program attendance, while the IV estimate is determined by the partial association between probability of leaving and the component of top program attendance correlated with the instrument – average past top program attendance. The results therefore mean that the association of propensity to leave with the component of top program attendance uncorrelated with the average past top program attendance is much smaller than the component that is correlated. This indicates that individuals who are selected into top programs either due to positive information spillovers in doctoral admissions processes or network effects are more likely to emigrate following graduation. The following scenario supports these findings – the students who would be differentially admitted into doctoral programs due to better information from past student history are most likely those who are in the bottom of the *within-program* skill distribution. Upon graduation, these are the students that are more likely to emigrate as compared to their higher skilled batchmates, since the U.S. labor market wouldn't place as much value on their skill but a top program degree still provides a lot of leverage in international labor markets. Thus, even though there is positive skill selection across the entire gamut of programs (skills), the results may be primarily driven by negative selection within top programs. This scenario also helps explain why the magnitude of skill selection is so large in the case of students from low/middle income countries, since a top program degree has much greater purchase in labor markets in low/middle income countries as compared to in high income countries. Unfortunately, the data provides no way for us to verify this narrative empirically.

The second possibility is that the instrument itself is endogenous and hence the IV results are biased upwards. We do not believe this to be the case since the instrument varies at the country of origin and year of graduation level, and we control for unobserved effects at that level through fixed effects. In sum, the IV results confirm the patterns in positive skill selection in the out-migration of foreign-born U.S. doctorates as uncovered in Chapter 1 and finds that it is of a much larger magnitude than naive OLS regressions would have us believe.

Concluding Remarks

This chapter investigates the relationship between attending a top program and the propensity to emigrate for foreign born doctoral students in the U.S. from a causal perspective. In this respect it seeks to build on the findings of the first chapter of this thesis, which finds that there is evidence of positive skill selection in the out-migration of these individuals and that the effect is driven entirely by the migration behavior of students from low/middle income countries. A possible caveat in of the analysis in the first chapter is that there may be many unobserved factors at the individual level which are correlated with both top program attendance and propensity to leave and may cause naive OLS results to be biased. In the presence of many such factors, the direction of the bias is difficult to pin down. This chapter aims resolves the issue of uncovering the true magnitude of skill selection in the out-migration of foreign-born doctoral students by using an instrumental variables approach.

We propose to instrument top program attendance (the measure of skill of an individual) using the past three year (from year of graduation) average of top program attendance from a student's country of origin. The reason we expect these variables to be associated is as follows. Firstly, the presence of doctoral students from a particular country of origin allows schools to elicit more information about the quality of students from that country. A larger number of students in the program, then, indicates that this information may be inducing schools to accept more students from the country in question and raises the probability of top program attendance at the individual level. Secondly, for any individual looking to apply to doctoral programs in the U.S., a larger presence of doctoral students from the individuals' country of origin in a top program may induce the individual to apply to that program.

Using this instrument, we find that although the OLS regressions in the first chapter identify the patterns of positive skill selection in out-migration of foreign-born doctoral students correctly, it severely underestimates the effects. The analysis in the this chapter verifies that there is indeed a strong causal relationship between attending a top program and leaving the U.S. following graduation, and this effect is entirely driven by students coming from low/middle income countries. These results bolster the narrative in the first chapter and indicate that there may indeed be some evidence to support the claim that the U.S. is losing top talent to global competitors, especially rapidly expanding low/middle income countries.

Tables

Table 3.1: First Stage Regressions: Full Sample and S&E Only

	All Fields	S&E Only
Dependent Variable: Attended a Top Program		
Average Past Top Program Attendance	.00277***	.00281***
	(.00087)	(.00088)
Bachelors' in the US	.01936	.00064
	(.03025)	(.02819)
Either Parent has Bachelors'	.04494**	.03592**
	(.01962)	(.01710)
Male	.06965***	.06490***
	(.01420)	(.01620)
Married	–.03239***	–.04545***
	(.01650)	(.01635)
Age	–.01355***	–.01405***
	(.00155)	(.00173)
US Permanent Resident	.00309	.00756
	(.02487)	(.02844)
US Citizen	.00101	.02618
	(.02788)	(.02942)
Received RA/TA	.06310***	.08218***
	(.01986)	(.02548)
Received Fellowship	.11962***	.09582**
	(.02791)	(.03755)
Received Foreign Support	.11550***	.13578***
	(.04236)	(.04636)
Relative GDP Growth	.00018	.00018
	(.00012)	(.00012)
Relative Unemployment	–.02436	–.03319
	(.01912)	(.02281)
FDI Inflows to Destination Country	.00331*	.00473**
	(.00175)	(.00195)
Relative Patenting Intensity	–.00003	–.00005
	(.00002)	(.00003)
Relative Rule of Law	–.00001	–.00008
	(.00018)	(.00015)
No. of Observations	6169	5238
Robust F Statistic	10.0463	10.12

Table 3.2: Determinants of Leaving the U.S: OLS vs. IV for Full Sample and S&E Only

	All Fields		S&E Only	
	OLS	IV	OLS	IV
Measured at time of PhD Receipt				
Attended a Top Program	.01416	.59639	.01612	.72474
	(.01320)	(.48036)	(.01255)	(.48563)
Bachelors' in the US	−.11983***	−.13159***	−.11448***	−.11560***
	(.02325)	(.02464)	(.02343)	(.02503)
Either Parent has Bachelors'	−.00479	−.02994*	−.00037	−.02451
	(.00830)	(.01702)	(.00783)	(.01545)
Male	.02281	−.01736	.01926	−.02637
	(.01622)	(.03913)	(.01556)	(.03743)
Married	−.04813*	−.02889*	−.05022**	−.01751
	(.02522)	(.01646)	(.02173)	(.02265)
Age	.00325**	.01121	.00210*	.01214
	(.00150)	(.00705)	(.00127)	(.00741)
US Permanent Resident	−.23822***	−.23842***	−.21432***	−.21754***
	(.05111)	(.04301)	(.05001)	(.04037)
US Citizen	−.24660***	−.24719***	−.23371***	−.25240***
	(.03441)	(.02988)	(.03414)	(.03316)
Received RA/TA	−.04342**	−.07918**	−.02571	−.08281*
	(.02001)	(.03666)	(.02197)	(.04776)
Received Fellowship	.01808	−.05037	.02961	−.03653
	(.02377)	(.06668)	(.02849)	(.06477)
Received Foreign Support	.12311***	.05524	.14519***	.04813
	(.03385)	(.07432)	(.03525)	(.08501)
Measured at time of Emigration (US Relative to Destination Country)				
Relative GDP Growth	−.00031***	−.00041***	−.00026***	−.00038**
	(.00010)	(.00011)	(.00008)	(.00015)
Relative Unemployment	.04939	.06297**	.04784	.07068**
	(.03479)	(.02874)	(.03419)	(.02995)
FDI Inflows to Destination Country	.01480***	.01298***	.01550***	.01229***
	(.00398)	(.00351)	(.00383)	(.00416)
Relative Patenting Intensity	−.00002***	−.00002**	−.00002**	−.00002*
	(.00001)	(.00001)	(.00001)	(.00001)
Relative Rule of Law	−.00015	−.00015	−.00024	−.00030
	(.00025)	(.00032)	(.00026)	(.00030)
No. of Observations	6169	6169	5238	5238

Table 3.3: Determinants of Leaving the U.S. by Country of Origin: OLS vs. IV

	From High Income		From Low/Middle Income	
	OLS	IV	OLS	IV
Measured at time of PhD Receipt				
Attended a Top Program	–.01542	–.53005*	.03919***	1.43301**
	(.02038)	(.29572)	(.00658)	(.58381)
Bachelors' in the US	–.13126***	–.14332***	–.07611***	–.20811**
	(.02735)	(.03306)	(.02689)	(.08659)
Either Parent has Bachelors'	–.01323	.02158	–.00385	–.01801
	(.01358)	(.02365)	(.00722)	(.03561)
Male	.05546**	.09928**	–.00352	–.06966
	(.02663)	(.04573)	(.00771)	(.04346)
Married	–.00894	–.02201	–.08665***	–.02696
	(.01937)	(.02755)	(.02746)	(.03334)
Age	.00190	–.00666	.00406**	.01882***
	(.00210)	(.00509)	(.00166)	(.00511)
US Permanent Resident	–.33534***	–.34480***	–.14183***	–.16422***
	(.02122)	(.02804)	(.03627)	(.03802)
US Citizen	–.27142***	–.28011***	–.18797***	–.26079***
	(.02820)	(.03052)	(.05237)	(.06153)
Received RA/TA	–.05507**	–.04125	–.02318	–.15767**
	(.02421)	(.02750)	(.02665)	(.06654)
Received Fellowship	.02302	.06524	.01176	–.19409**
	(.03571)	(.04299)	(.01805)	(.09208)
Received Foreign Support	.12534***	.18321***	.09177**	–.03370
	(.03975)	(.04337)	(.04162)	(.05698)
Measured at time of Emigration (US Relative to Destination Country)				
Relative GDP Growth	–.00038***	–.00024**	.00089	.00274*
	(.00009)	(.00012)	(.00073)	(.00149)
Relative Unemployment	.19805**	.19159**	–.02316	.00366
	(.08524)	(.08125)	(.02797)	(.05088)
FDI Inflows to Destination Country	.00860**	.00954**	.02459***	.01587**
	(.00422)	(.00430)	(.00493)	(.00756)
Relative Patenting Intensity	–.00048	–.00065**	–.00002***	–.00002*
	(.00029)	(.00031)	(.00001)	(.00001)
Relative Rule of Law	–.00006	–.00007	–.00023	–.00018
	(.00021)	(.00011)	(.00049)	(.00043)
No. of Observations	3111	3111	3058	3058
First Stage: Average Past Top Program Attendance		.01079**		.00275**
		(.00428)		(.00129)

Table 3.4: Determinants of Leaving the U.S. by Country of Origin for S&E Only: OLS vs. IV

	From High Income		From Low/Middle Income	
	OLS	IV	OLS	IV
Measured at time of PhD Receipt				
Attended a Top Program	–.00809	–.53877*	.03411***	1.33926***
	(.02386)	(.28989)	(.01056)	(.47485)
Bachelors' in the US	–.13991***	–.15563***	–.05350*	–.13083*
	(.02750)	(.03395)	(.02986)	(.07502)
Either Parent has Bachelors'	–.00399	.02116	–.00336	–.02222
	(.01431)	(.01894)	(.00612)	(.03698)
Male	.05090*	.09355**	–.00452	–.06573
	(.02706)	(.04453)	(.00961)	(.04243)
Married	–.03283*	–.05459*	–.06188**	.00464
	(.01922)	(.03029)	(.02983)	(.03256)
Age	.00143	–.00763	.00206	.01781***
	(.00212)	(.00593)	(.00135)	(.00506)
US Permanent Resident	–.30781***	–.31995***	–.12964***	–.16338***
	(.02761)	(.03182)	(.03479)	(.03408)
US Citizen	–.24564***	–.24532***	–.19833***	–.29937***
	(.02642)	(.02737)	(.06179)	(.07542)
Received RA/TA	–.03850	–.01612	–.00381	–.15143**
	(.02996)	(.03303)	(.02699)	(.06305)
Received Fellowship	.03948	.06409	.02494	–.14860**
	(.04361)	(.05145)	(.01891)	(.06830)
Received Foreign Support	.14201***	.20658***	.14197***	–.03207
	(.03964)	(.04813)	(.05171)	(.06472)
Measured at time of Emigration (US Relative to Destination Country)				
Relative GDP Growth	–.00033****	–.00021**	.00108	.00286**
	(.00005)	(.00009)	(.00076)	(.00138)
Relative Unemployment	.18949**	.17252**	–.02099	.01224
	(.08392)	(.08225)	(.02865)	(.05633)
FDI Inflows to Destination Country	.00975**	.01176**	.02284***	.01420*
	(.00444)	(.00475)	(.00510)	(.00858)
Relative Patenting Intensity	–.00047	–.00058*	–.00002*	–.00001
	(.00031)	(.00030)	(.00001)	(.00001)
Relative Rule of Law	–.00013	–.00008	–.00031	–.00035
	(.00018)	(.00010)	(.00051)	(.00045)
No. of Observations	2476	2762	2476	2762
First Stage: Average Past Top Program Attendance		.01160***		.00326***
		(.00370)		(.00130)

REFERENCES

Banerjee, Abhijit V, and Esther Duflo. 2005. Growth theory through the lens of development economics. *Handbook of economic growth* 1:473–552.

Bartelsman, Eric J., John C. Haltiwanger, and Stefano Scarpetta. 2009. Cross-country differences in productivity: The role of allocation and selection. Tech. Rep.

Becker, Gary S. 1973. A theory of marriage: Part i. *Journal of Political economy* 81(4):813–846.

Bernanke, Ben S, Jean Boivin, and Piotr Eliasz. 2005. Measuring the effects of monetary policy: a factor-augmented vector autoregressive (favar) approach. *The Quarterly journal of economics* 120(1):387–422.

Black, Grant, and Paula Stephan. 2007. The importance of foreign PhD students to US science. *Science and the University* 113–33.

Borjas, George J. 1989. Immigrant and emigrant earnings: A longitudinal study. *Economic inquiry* 27(1):21–37.

———. 2005. The labor-market impact of high-skill immigration. *American Economic Review* 95(2):56–60.

Borjas, George J., and Bernt Bratsberg. 1996. Who leaves? the outmigration of the foreign-born. *The Review of Economics and Statistics* 78(1):165–176.

Bound, John, Sarah Turner, and Patrick Walsh. 2009. Internationalization of US doctorate education. Tech. Rep., National Bureau of Economic Research.

Burns, Arthur F, Wesley C Mitchell, et al. 1946. Measuring business cycles. *Nber Books*.

Chellaraj, Gnanaraj, Keith E Maskus, and Aaditya Mattoo. 2008. The contribution of international graduate students to US innovation. *Review of International Economics* 16(3):444–462.

Chiappori, Pierre-André, and Bernard Salanié. 2016. The econometrics of matching models. *Journal of Economic Literature* 54(3):832–861.

Choo, Eugene, and Aloysius Siow. 2006. Who marries whom and why. *Journal of political Economy* 114(1):175–201.

Constant, Amelie, and Douglas S. Massey. 2003. Self-selection, earnings, and out-migration: A longitudinal study of immigrants to germany. *Journal of Population Economics* 16(4):631–653.

Edin, Per-Anders, Robert J LaLonde, and Olof Åslund. 2000. Emigration of immigrants and measures of immigrant assimilation: Evidence from sweden. Tech. Rep., Working Paper, Department of Economics, Uppsala University.

Ehrenberg, Ronald G. 2012. American higher education in transition. *Journal of Economic Perspectives* 26(1):193–216.

Finn, Michael G. 2010. Stay rates of foreign doctorate recipients from US universities, 2007.

———. 2014. Stay rates of foreign doctorate recipients from US universities, 2011.

Freeman, Richard B. 2010. What does global expansion of higher education mean for the united states? In *American universities in a global market*, 373–404. University of Chicago Press.

Galichon, Alfred, and Bernard Salanié. 2015. Cupids invisible hand: Social surplus and identification in matching models.

Gang, Ira N, and Thomas K Bauer. 1998. Temporary migrants from egypt: How long do they stay abroad?

Graham, Bryan. 2011. Econometric methods for the analysis of assignment problems in the presence of complementarity and social spillovers. *Handbook of social economics* 1:965–1052.

Grogger, Jeffrey, and Gordon H Hanson. 2015. Attracting talent: Location choices of foreign-born PhDs in the united states. *Journal of Labor Economics* 33(S1 Part 2):S5–S38.

Hagedorn, Marcus, Tzuo Hann Law, and Iourii Manovskii. 2017. Identifying equilibrium models of labor market sorting. *Econometrica* 85(1):29–65.

Hsieh, Chang-Tai, and Peter J. Klenow. 2009. Misallocation and manufacturing tfp in china and india. *The Quarterly Journal of Economics* 124(4):1403–1448.

Hunt, Jennifer. 2011. Which immigrants are most innovative and entrepreneurial? distinctions by entry visa. *Journal of Labor Economics* 29(3):417–457.

Hunt, Jennifer, and Marjolaine Gauthier-Loiselle. 2008. How much does immigration boost innovation? Tech. Rep., National Bureau of Economic Research.

Jasso, Guillermina, and Mark R Rosenzweig. 1988. How well do us immigrants do? vintage effects emigration selectivity and occupational mobility. *Research in Population Economics* 6.

Jeong, Hyeok, and Robert M Townsend. 2007. Sources of tfp growth: occupational choice and financial deepening. *Economic Theory* 32(1):179–221.

Kapur, Devesh, and John McHale. 2005. *Give us your best and brightest: The global hunt for talent and its impact on the developing world*. Center for Global Development Washington, DC.

Kim, Chang-Jin, Charles R Nelson, et al. 1999. State-space models with regime switching: classical and gibbs-sampling approaches with applications. *MIT Press Books* 1.

Levin, Sharon G, and Paula E Stephan. 1999. Are the foreign born a source of strength for US science? *Science* 285(5431):1213–1214.

MCFADDEN, D. 1974. Conditional logit analysis of qualitative choice behavior. *Frontiers in Econometrics* 105–142.

National Academy of Sciences and National Academy of Engineering and Institute of Medicine. 2007. Rising above the gathering storm: Energizing and employing america for a brighter economic future.

National Science Board (NSB). 2014. Science and engineering indicators 2014.

Parsons, Christopher. 2015. Whats the best policy to attract high-skilled migrants? *World Ecocnomic Forum Global Agenda*.

Peri, Giovanni. 2007. Higher education, innovation and growth.

Reagan, Patricia B, and Randall J Olsen. 2000. You can go home again: Evidence from longitudinal data. *Demography* 37(3):339–350.

Restuccia, Diego, and Richard Rogerson. 2008. Policy distortions and aggregate productivity with heterogeneous establishments. *Review of Economic Dynamics* 11(4):707–720.

Rothwell, Jonathan. 2014. Still searching: Job vacancies and stem skills. *Metropolitan Policy Program, Brookings Institution*.

Sargent, Thomas J, Christopher A Sims, et al. 1977. Business cycle modeling without pretending to have too much a priori economic theory. *New methods in business cycle research* 1:145–168.

Shimer, Robert, and Lones Smith. 2000. Assortative matching and search. *Econometrica* 68(2):343–369.

Staiger, Douglas, James H Stock, et al. 1997. Instrumental variables regression with weak instruments. *Econometrica* 65(3):557–586.

Stock, James, and Mark W. Watson. 1991. *A probability model of the coincident economic indicators*, 63–90. Cambridge University Press.

Stock, James, and Motohiro Yogo. 2005. *Testing for weak instruments in linear iv regression*, 80–108. New York: Cambridge University Press.

Stock, James H, and Mark Watson. 2011. Dynamic factor models. *Oxford Handbook on Economic Forecasting*.

Stock, James H, and Mark W Watson. 2002. Macroeconomic forecasting using diffusion indexes. *Journal of Business & Economic Statistics* 20(2):147–162.

Stuen, Eric T, Ahmed Mushfiq Mobarak, and Keith E Maskus. 2012. Skilled immigration and innovation: evidence from enrollment fluctuations in US doctoral programmes. *The Economic Journal* 122(565):1143–1176.

US Congress Joint Economic Committee and others. 2012. Stem education: Preparing for the jobs of the future.

Xu, Rui. 2015. High-skilled migration and global innovation.

A

APPENDIX A: VARIABLE NAMES AND DATA SOURCES (CHAPTER 1)

Table A.1: Variable Descriptions and Data Sources

Variable	Description	Source
Dependent Variable		
Left	Indicator for whether individual left the U.S.	SDR
Explanatory Variables		
Attended a Top Program	Indicator for whether individual attended a top program within his/her field. Top programs are identified according to Finn (2010).	SED
Bachelors' in the US	Indicator for whether individual has a U.S. undergraduate degree.	SED
Either Parent has Bachelors'	Indicator for whether either parent of the individual attained a bachelor's degree	SED
Male	Indicator for gender of individual	SED
Married	Indicator for whether an individual is presently married.	SDR
Age	Age at Ph.D. completion date.	SED
US Permanent Resident	Indicator for whether the individual has/ever had permanent residence status in the U.S.	SED
US Citizen	Indicator for whether the individual is a U.S. citizen	SED
Received RA/TA	Indicator for whether the individual received RA/TA during Ph.D.	SED
Received Fellowship	Indicator for whether the individual received a Fellowship for Ph.D. studies	SED
Received Foreign Support	Indicator for whether the individual received support from home country for Ph.D. studies	SED
Relative GDP Growth	Ratio of standardized US per-capita GDP growth rate averaged over three years prior to individual's graduation year to the standardized Home/Work country per-capita GDP growth rate averaged over three years prior to individual's graduation year.	WDI
Relative Unemployment	Ratio of US unemployment rate (ILO measure) averaged over three years prior to individual's graduation year to the Home/Work country unemployment rate (ILO measure) averaged over three years prior to individual's graduation year.	WDI
FDI Inflows to Destination Country	Net FDI inflows to Home/Work Country (in constant 2005 $) averaged over three years prior to individual's graduation year.	WDI
Relative Patenting Intensity	Ratio of US patents (resident) per-capita averaged over three years prior to individual's graduation year to the Home/Work country patents (resident) per-capita averaged over three years prior to individual's graduation year.	WIPO
Relative Rule of Law	Ratio of US estimated rule of law to the Home/Work country estimated rule of law	WDI
Relative Political Stability	Ratio of US political stability (precentile) to the Home/Work country political stability (percentile)	WDI
Job in Field in which Trained	Indicator for whether work on principal job was "closely" related to Ph.D. field (self reported)	SDR
Salary Premium	Log difference between average salary for an individual's job type in Home/Work country and the average salary for an individual's job type in the US	SDR

B

APPENDIX B: DEGREE FIELD GROUPINGS (CHAPTER 2)

1 - Agricultural Sciences

Agricultural Economics

Agricultural Business & Management

Agricultural Animal Breeding

Animal Husbandry

Animal Nutrition

Dairy Science

Animal Science, Poultry (or Avian)

Animal Science, Other

Agronomy & Crop Science

Agricultural & Horticultural Plant Breeding

Plant Pathology/Phytopathology

Plant Protect/Pest Management

Plant Sciences, Other

Food Sciences

Food Distribution

Food Science

Food Sciences and Technology, Other

Soil Sciences

Soil Chemistry/Microbiology

Soil Sciences, Other

Horticulture Science

Fish & Wildlife

Fishing and Fisheries Sciences/Management

Wildlife Management

Forestry Science

Forest Sciences and Biology

Forest Engineering

Forest/Resources Management

Wood Science & Pulp/Paper Technology

Natural Resources/Conservation

Forestry & Related Science, Other

Wildlife/Range Management

Environmental Science

Agriculture, General

Agricultural Science, Other

2 - Biological Sciences

Biochemistry

Bioinformatics

Biomedical Sciences

Biophysics

Biotechnology

Bacteriology

Plant Genetics

Plant Pathology/Phytopathology

Plant Physiology

Botany/Plant Biology

Anatomy

Biometrics & Biostatistic

Cell/Cellular Biology and Histology

Evolutionary Biology

Ecology

Hydrobiology

Developmental Biology/Embryology

Endocrinology

Entomology

Immunology

Molecular Biology

Microbiology & Bacteriology

Microbiology

Cancer Biology

Neurosciences
Nutrition Sciences
Parasitology
Toxicology
Genetics/genomics, Human & Animal
Genetics
Pathology, Human & Animal
Pharmacology, Human & Animal
Physiology, Human & Animal
Animal & Plant Physiology
Zoology
Biology/Biomedical Sciences, General
Biology/Biomedical Sciences, Other

3 - Health Sciences

Speech-Language Pathology & Audiology
Dentistry
Environmental Health
Environmental Toxicology
Health Systems/Services Administration
Public Health
Public Health & Epidemiology
Epidemiology
Kinesiology/Exercise Science
Hospital Administration
Medicine & Surgery
Nursing Science
Optometry & Ophthalmology
Medicinal/Pharmaceutical Sciences
Rehabilitation/Therapeutic Services
Veterinary Sciences
Health Sciences, General
Health Sciences, Other

4 - Engineering

Aerospace, Aeronautical & Astronautical
Agricultural
Bioengineering & Biomedical
Ceramic Sciences
Chemical
Civil
Communications
Electrical
Electronics
Electrical, Electronics and Communications
Engineering Mechanics
Engineering Physics
Engineering Science
Environmental Health Engineering
Industrial & Manufacturing
Materials Science
Mechanical
Metallurgical
Mining & Mineral
Naval Architecture & Marine Engineering
Nuclear
Ocean
Operations Research
Petroleum
Polymer & Plastics
Systems
Textile

Engineering Management & Administration
Engineering, General
Engineering, Other

5 - Computer Sciences and Engineering
Computer Enigeering
Computer Science
Information Science & Systems
Computer & Information Science, Other

6 - Mathematics
Applied Mathematics
Algebra
Analysis & Functional Analysis
Geometry/Geometric Analysis
Logic
Number Theory
Statistics
Topology/Foundations
Computing Theory & Practice
Operations Research
Mathematics/Statistics, General
Mathematics/Statistics, Other

7 - Chemistry
Analytical Chemistry
Agriculture & Food Chemistry
Inorganic Chemistry
Nuclear Chemistry
Organic Chemistry
Medicinal/Pharmaceutical Chemistry
Physical Chemistry
Polymer Chemistry
Theoretical Chemistry
Chemistry, General
Chemistry, Other

8 - Geological and Related Sciences
Geology
Geochemistry
Geophysics & Seismology
Geophysics (solid earth)
Paleontology
Fuel Technology & Petroleum Engineering
Mineralogy & Petrology
Mineralogy/Petrology/Geological Chemistry
Stratigraphy & Sedimentation
Geomorphology & Glacial Geology
Applied Geology
Applied Geology/Geological Engineering
Geological and Earth Sciences, General
Geological and Earth Sciences, Other

9 - Physics
Acoustics
Atomic/Molecular/Chemical Physics
Electron Physics
Electromagnetism
Particle (Elementary) Physics
Biophysics
Fluids Physics
Mechanics

Nuclear Physics
Optics/Phototonics
Plasma/Fusion Physics
Polymer Physics
Thermal Physics
Condensed Matter/Low Temperature Physics
Theoretical Physics
Applied Physics
Physics, General
Physics, Other

10 - Other Physical Sciences

Astronomy
Astrophysics
Astronomy & Astrophysics
Atmospheric Chemistry and Climatology
Atmospheric Physics and Dynamics
Meteorology
Atmospheric Science/Meteorology, General
Atmospheric Science/Meteorology, Other
Environmental Science
Hydrology & Water Resources
Oceanography, Chemical and Physical
Marine Sciences
Ocean/Marine, Other

11 - Psychology

Clinical Psychology
Cognitive Psychology & Psycholinguistics
Comparative Psychology
Counseling
Developmental & Child Psychology
Human Development & Family Studies
Experimental Psychology
Experimental/Comparative & Physiological Psychology
Educational Psychology
Human Engineering
Family Psychology
Industrial & Organizational Psychology
Personality Psychology
Physiological/Psychobiology
Psychometrics
Psychometrics and Quantitative Psychology
School Psychology
Social Psychology
Psychology, General
Psychology, Other

12 - Economics

Economics
Econometrics
Public Policy Analysis
Statistics

C

APPENDIX C: JOB CATEGORY CLASSIFICATIONS (CHAPTER 2)

Table C.1: Job Category Classifications

	Code	Description	Jobtype	Jobtype1	Jobtype2
1	110510	Computer and information scientists, research	Computer Sciences and Engineering		
2	110530	Computer support specialists	Computer Sciences and Engineering		
3	110540	Computer system analysts	Computer Sciences and Engineering		
4	110550	Database administrators	Computer Sciences and Engineering		
5	110560	Network and computer systems administrators	Computer Sciences and Engineering		
6	110570	Network systems and data communications analysts	Computer Sciences and Engineering		
7	110580	OTHER computer information science occupations	Computer Sciences and Engineering		
8	110880	Computer engineers - software	Computer Sciences and Engineering		
9	121720	Mathematicians	Mathematics		
10	121730	Operations research analysts, including modeling	Mathematics	Engineering	
11	121740	Statisticians	Mathematics	Economics	
12	121760	OTHER mathematical scientists	Mathematics		
13	182760	Postsecondary Teachers: Computer Science teachers	Computer Sciences and Engineering		
14	182860	Postsecondary Teachers: Mathematics and statistics t...	Mathematics		
15	210210	Agricultural and food scientists	Agricultural Sciences		
16	220220	Biochemists and biophysicists	Biological Sciences		
17	220230	Biological scientists [e.g., botanists, ecologists,...	Biological Sciences		
18	220250	Medical scientists [excluding practitioners]	Health Sciences		
19	220270	OTHER biological and life scientists	Biological Sciences		
20	230240	Forestry and conservation scientists	Agricultural Sciences		

	Code	Description	Jobtype	Jobtype1	Jobtype2
21	282710	Postsecondary Teachers: Agriculture teachers	Agricultural Sciences		
22	282730	Postsecondary Teachers: Biological Sciences teachers...	Biological Sciences		
23	282970	Postsecondary Teachers: OTHER natural sciences teach...	Agricultural Sciences	Biological Sciences	Health Sciences
24	311930	Chemists, except biochemists	Chemistry		
25	321920	Atmospheric and space scientists	Other Physical Sciences		
26	321940	Geologists, including earth scientists	Geological Sciences		
27	321950	Oceanographers	Other Physical Sciences		
28	331910	Astronomers	Other Physical Sciences		
29	331960	Physicists	Physics		
30	341980	OTHER physical scientists	Other Physical Sciences		
31	382750	Postsecondary Teachers: Chemistry	Chemistry		
32	382770	Postsecondary Teachers: Earth, Environmental	Other Physical Sciences		
33	382890	Postsecondary Teachers: Physics	Physics		
34	412320	Economists	Economics		
35	432360	Psychologists, including clinical	Psychology		
36	482780	Postsecondary Teachers: Economics	Economics		
37	482910	Postsecondary Teachers: Psychology	Psychology		
38	510820	Aeronautical, aerospace, astronautical engineers	Engineering		
39	520850	Chemical engineers	Engineering		
40	530860	Civil, including architectural and sanitary engineer	Engineering		
41	540870	Computer engineer - hardware	Computer Sciences and Engineering		
42	540890	Electrical and electronics engineers	Engineering		
43	550910	Industrial engineers	Engineering		
44	560940	Mechanical engineers	Engineering		
45	570830	Agricultural engineers	Engineering		

	Code	Description	Jobtype	Jobtype1	Jobtype2
46	570840	Bioengineering and biomedical engineers	Engineering		
47	570900	Environmental engineers	Engineering		
48	570920	Marine engineers and naval architects	Engineering		
49	570930	Materials and metallurgical engineers	Engineering		
50	570950	Mining and geological engineers	Engineering		
51	570960	Nuclear engineers	Engineering		
52	570970	Petroleum engineers	Engineering		
53	570980	Sales engineers	Engineering		
54	570990	Other engineers	Engineering		
55	582800	Postsecondary Teachers: Engineering	Engineering		
56	611110	Diagnosing/Treating Practitioners	Health Sciences		
57	611120	RNs, pharmacists, dieticians, therapists	Health Sciences		
58	611130	Health Technologists and Technicians	Health Sciences		
59	611140	OTHER health occupations	Health Sciences		
60	612870	Postsecondary teachers - Health and related sciences	Health Sciences		
61	640260	Technologists/technicians in biological/life	Biological Sciences		
62	640520	Computer programmers	Computer Sciences and Engineering		
63	641000	Electrical, electronic, industrial, mechanical technicians	Engineering		
64	641010	Drafting occupations, including computer drafting	Computer Sciences and Engineering		
65	641030	OTHER engineering technologists and technicians	Engineering		
66	641970	Technologists and technicians in the physical sciences	Engineering		
67	721520	Personnel, training, and labor relations specialists	Psychology		
68	750700	Counselors, Educational, vocational, mental health	Psychology		

CURRICULUM VITA

CHANDRAMOULI BANERJEE

EDUCATION

- University of Wisconsin - Milwaukee. 2013 - Present

  Ph.D., Economics

  Dissertation: "Essays on the Career Choices of Doctoral Students in the U.S."

- Hindu College, University of Delhi, New Delhi, India. 2009 - 2012

  Bachelor of Arts (Honours), Economics

WORKING PAPERS

- Job-Skill Match in the Labor Market for Scientists and its Aggregate Implications.

- Explaining the Effect of Financial Development on Property Rights (with Niloy Bose and Chitralekha Rath).

- To Stay or Not to Stay: Location Choice of Foreign-born U.S. PhDs (with Scott J. Adams and Niloy Bose).

WORK IN PROGRESS

- Revisiting Location Choice - A Structural Approach.

- Skill Selection in the Out-Migration of the Highly Skilled Foreign-born: A Causal Approach (with Scott J. Adams and Niloy Bose).

- The Impact of Training on Loan Delinquency (with Niloy Bose and Tianfang Li)

TEACHING EXPERIENCE

- Department of Economics, University of Wisconsin - Milwaukee
  Instructor 2015 - Present
  - Economic Statistics: Summer 2016, Fall 2016
  - Principles of Macroeconomics: Fall 2015, Spring 2016, Spring 2017, Fall 2017.
- Department of Economics, University of Wisconsin - Milwaukee
  Graduate Teaching Assistant 2013 - Present
  - Principles of Macroeconomics: Fall 2013-2017, Spring 2014-2017.
  - Econometrics (M.A. Economics Program): Fall 2014
  - Microeconomics (M.A. Economics Program): Fall 2014
  - Macroeconomics (M.A. Economics Program): Spring 2015

CONFERENCE PRESENTATIONS

- Applied Microeconomics Seminar Series
  Department of Economics, University of Wisconsin - Milwaukee October 2017, January 2016
- Midwest Economics Association Conference
  Cincinnati OH March, 2017

- 12th Annual Conference on Economic Growth and Development
  Indian Statistical Institute, New Delhi, India December, 2016

SELECTED AWARDS AND RECOGNITIONS

- J. Walter Elliot Award for Excellence in Macroeconomics, University of Wisconsin - Milwaukee 2014
- Chancellor's Graduate Student Award, University of Wisconsin - Milwaukee 2017, 2014 - 2015, 2013 - 2014
- Graduate School Travel Award, University of Wisconsin - Milwaukee 2016