The Impact of International Graduate Students on University Innovation

by

Enver Ferit Akın

A Dissertation Submitted to the
Graduate School of Social Sciences and Humanities
in Partial Fulfillment of the Requirements for
the Degree of

Master of Arts

in

Economics

KOÇ ÜNİVERSİTESİ

August 12, 2024

The Impact of International Graduate Students on University Innovation

Koç University
Graduate School of Social Sciences and Humanities
This is to certify that I have examined this copy of a master's thesis by

Enver Ferit Akın

and have found that it is complete and satisfactory in all respects,
and that any and all revisions required by the final
examining committee have been made.

Committee Members:

____

Asst. Prof. Murat Demirci (Advisor)

____

Assoc. Prof. Emre Ekinci (Co-Advisor)

____

Prof. Seyit Mümin Cilasun

____

Asst. Prof. Ömer Faruk Sözbir

Date: 12.08.2024

ABSTRACT

The Impact of International Graduate Students on University Innovation

Enver Ferit Akın

Master of Arts in Economics

August 12, 2024

This thesis explores the impact of international graduate students in Science, Technology, Engineering, and Mathematics (STEM) fields in the United States on the innovative outputs of U.S. universities. The analysis focuses on both innovation production of universities, measured by the patent grants, and the quality of these innovations, measured by the citations received by these patents. The study examines biennial periods from 1994 to 2016. To capture the causality, I employ the instrumental variables (IV) approach. The instrument exploits the historical distribution of international graduate students as a predictor of future migration flows. The findings indicate a positive and significant impact of international graduate students on patent counts, but no significant effect on patent citations.

ÖZETÇE

Uluslarası Lisansüstü Öğrencilerin Üniversite İnovasyonuna Etkisi

Enver Ferit Akın

Ekonomi, Yüksek Lisans

12 Ağustos 2024

Bu tez, Amerika Birleşik Devletleri'nde Fen, Teknoloji, Mühendislik ve Matematik alanlarındaki uluslararası lisansüstü öğrencilerin, ABD üniversitelerinin inovatif çıktıları üzerindeki etkisini incelemektedir. Analiz, hem üniversitelerin inovasyon üretimini (patent sayıları ile ölçülen) hem de bu inovasyonların kalitesini (bu patentlerin aldığı atıflarla ölçülen) ele almaktadır. Çalışma, 1994'ten 2016'ya kadar olan iki yıllık dönemleri incelemektedir. Nedenselliği yakalamak için Araç Değişkeni (AD) yaklaşımı kullanılmaktadır. Araç değişkeni, gelecekteki göç akışlarının bir öngörücüsü olarak uluslararası lisansüstü öğrencilerin tarihsel dağılımından yararlanmaktadır. Bulgular, uluslararası lisansüstü öğrencilerin patent sayıları üzerinde pozitif ve anlamlı bir etkisi olduğunu, ancak patent atıfları üzerinde anlamlı bir etkisi olmadığını göstermektedir.

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to my advisors Assist. Prof. Murat Demirci and Assoc. Prof. Emre Ekinci for their invaluable support and encouragement throughout this difficult journey. Their wisdom, patience, and encouragement have been my guiding lights through this challenging yet rewarding journey. Additionally, I am grateful to TÜBİTAK for providing generous financial support. I am also grateful to my friends - Özlem, Eymen, Oğuzhan, Vedat, Nil, Bilal, Atakan, Fırat, and Berke for their help and companionship. This journey would have been immeasurably more difficult and far less enriching without their presence and support. Finally, I would like to thank my mother and father, and my family members for their endless support and belief in me. Their unconditional love has been the greatest source of strength throughout this endeavor.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS

IV	Instrumental Variable
STEM	Science, Technology, Engineering, and Mathematics
US	United States
OLS	Ordinary Least Squares
CF	Control Function
FE	Fixed Effects
FENB	Fixed Effects Negative Binomial
USPTO	United States Patent and Trademark Office
R&D	Research and Development
IPC	International Patent Classification

Chapter 1:
INTRODUCTION

The impact of immigrants on the economy are multifaceted and can vary based on the socioeconomic factors of the host country. Broadly, immigration affects labor market outcomes, innovation and entrepreneurship, public finance, and economic growth. This thesis focuses on the innovation dimension. Since Schumpeter's seminal idea, "creative destruction", innovation has been regarded as one of the main drivers of economic growth. The crucial role of a skilled workforce in the production of innovation suggests that skilled immigration can play a critical role in this process. Consequently, both immigration and innovation are essential components of economic prosperity.

The United States, with its numerous prestigious universities and favorable socioeconomic conditions for graduate students, is a preferred destination for international graduate students.[1] Concurrently, the role of universities in technological improvements has been growing, as evidenced by the increasing number of patents granted to US universities (Hall, 2004). Therefore, universities serve as both magnets for skilled immigrants and as hubs of research and innovation. They provide ideal environments for analyzing the interplay between skilled migration and the innovation production capacity of the United States.

In this thesis, I investigate the impact of international graduate students in science, technology, engineering, and mathematics (STEM) fields on the innovative output of US universities.[2] The analysis focuses on both innovation production of universities, measured by the patent grants, and the quality of these innovations, measured by the citations received by those patents. The study examines biennial periods from 1994 to 2016.

To uncover the causal relationship between the innovative output of universities and international graduate students in STEM fields, I employ a shift-share (or "Bartik") instrumental variables (IV) approach. This entails combining a university's predetermined share of nationwide international graduate STEM enrollment with the

[1] According to the *Open Doors Report on International Educational Exchange*, there are 467,027 international graduate students enrolled in the US universities in the 2022/2023 education year.
[2] STEM fields include biological sciences/life sciences, mathematics, engineering, physical sciences, dentistry, and medicine.

national level of supply of international graduate STEM students. The predetermined share is calculated based on the year 1986. Using the IV approach and incorporating numerous fixed effects, time trends, and control variables (such as R&D spending and the number of domestic graduate students), the causal effect is estimated by employing both the 2SLS method and a fixed-effects Poisson with control function approach. The findings indicate a positive and significant impact of international graduate students on patent counts, but no significant effect on patent citations.

Additionally, various robustness checks are conducted. These include the utilization of an alternative instrument which leverages the fluctuations in enrollments across STEM majors, the use of a binary dependent variable, the use of a truncation adjusted dependent variable, and the use of the share of international graduate STEM students in the university as dependent variable. The results reveal that the effect on the number of patents is generally positive, but the statistical significance of the estimated impact disappears in most of these checks.

The rest of this thesis is structured as follows. Section II reviews the literature. Section III describes the data. Section IV outlines the methodology. Section V presents the results, and Section VI concludes.

Chapter 2:
LITERATURE REVIEW

There has been a lot of research conducted on how immigration affects host economies. For instance, Card (1990), Borjas (1994), and Borjas et al. (1997) provide seminal examples that looked at the impact of immigrant labor on native workers' earnings. These findings in these studies show that the degree of complementarity between native and immigrant workers determines how immigrants affect native workers' salaries.

This thesis focuses on two different key factors: foreign graduate students in the US as immigrant group of interest and the patenting activity of US universities as the outcome of interest. These factors have been altered many times by several policy changes over the last few decades. The rise in foreign student enrolment since the 1970s is largely due to relaxed immigration regulations. Notably, the Hart-Cellar Immigration Act of 1965 and the Kennedy-Rodino Immigration Act of 1990 accelerated the inflow of foreign students to the US (Chellaraj et al., 2008). Concurrently, the 1980 Bayh-Dole Act allowed U.S. colleges to market research findings, which fundamentally changed how they viewed intellectual property (Mowery et al., 2001).

The relationship between university research, innovation, and economic outcomes has been extensively studied, with notable contributions from Thursby and Kemp (2002), Jaffe and Trajtenberg (2005), Kim et al. (2007), and Babina et al. (2023). However, there is a notable gap in the literature regarding the specific impact of foreign graduate students on the innovative outputs of US universities. This thesis seeks to fill that gap by drawing from and expanding upon a few seminal works in this field.

Black et al. (2010) use PhD students, postdoctoral researchers, and professors as human resource inputs to estimate a knowledge production function for university patenting. Their results show that there is a positive and significant link between patent output and PhDs/Postdocs where one more PhD/Postdoc would raise patents by 10%. They find no significant relationship between temporary visa PhD students and patent output. In contrast to the current thesis, their analysis does not account for potential endogeneity and employs a different estimation technique (the individual effects negative binomial model).

Chellaraj et al. (2008) examine the impact of foreign graduate students on the patent applications and grants of both US universities and firms. They employ a modified version of “national ideas production function”. Their measures of new ideas production are total patent applications and patents granted to US universities and firms. They integrate international students and skilled immigrants as components of the inputs into the “idea production function”. They find that foreign graduate students have a positive and significant effect on patent applications grants. However, like Black et al., they do not address endogeneity concerns.

The work of Stuen et al. (2012) is particularly relevant to this thesis. In their study, spanning 1973-1998 and encompassing 2300 departments, they focus mainly on scientific publications and citations as proxies for innovation. Crucially, they address endogeneity concerns using an instrumental variable approach that exploits macroeconomic and policy shocks in source countries. Through an instrumental variable (IV) strategy that exploits macroeconomic and policy shocks in source countries, they capture exogenous variations of PhD student migration to the US. The rationale behind their IV approach is rooted in historical patterns of international student enrollment. They argue that the impact of student-supply shocks from source countries differs by field and university due to established preferences and acceptance patterns. Therefore, their university and field-specific instrument allows them to exploit variation across all dimensions of foreign PhD student count data. Formally, their instrument for a given field-university in a specific year is constructed as the product of three components: (1) the supply shock in a specific region and year, (2) the fraction of a university's foreign students from that region at an initial date, and (3) the fraction of students in a specific field from that region at the initial date. They find that both domestic and foreign PhD students increase scientific publications and citations in US science and engineering departments. Additionally, they demonstrate the importance of foreign PhD student quality in producing novel research through a theoretical model of admission decisions. While this thesis employs a similar identification strategy for causality, the focus is patenting (rather than publishing as in Stuen et al. (2012)) as the outcome of interest.

Chapter 3:
DATA

3.1 Patents and Citations

I create the counts of granted patent applications and patent citations associated with the United States universities for each year between 1975-2022 with the data from the *PatentsView,* a platform that focuses on intellectual property data. This platform is supported by the Office of the Chief Economists at the United States Patent and Trademark Office (USPTO). The *PatentsView* database has different datasets; each is about different aspects of patents, such as assignees, application information, and citations. First, I use the disambiguated assignee dataset containing information of granted patents and their assignees to construct a sample of university patents. The disambiguated assignee dataset covers 460,963 unique assignees and 7,334,517 granted patents as of December 2022.

Before moving forward, it is essential to elucidate several key terms related to patents. Patent application process consists of two primary stages: the application stage and the grant stage. I analyze the counts of ‘granted patent applications’ indicating that the applications of the patents have been successfully approved by the USPTO. I use the application date of the patents to determine the time of the innovation. While there are both granted and ungranted patents, I focus on granted patent applications to include innovative outputs that have received approval from the USPTO. A patent assignee is an organization or individual that possesses the property right to the patent. Typically, an assignee is the organization that employs the inventor of the technology. In our case, the patent assignees are universities. Disambiguation is needed for the following reason that is elaborated in the next paragraph.

One of the most critical problems of the patent assignee data is that the same assignee might be recorded with a different name in the data. For example, Stanford University is recorded as ‘Board of Trustees Leland Stanford Jr University’ and ‘Stanford University.’ Moreover, there are misspellings of the assignee names in the data. For instance, ‘Harvard University’ is recorded as ‘Harvard Universtyi’ for some patents. Also, some universities have independent technology transfer offices, such as The Wisconsin Alumni Research Foundation, which serves the University of Wisconsin-Madison. Also, in the assignee dataset, both the University of Wisconsin-Madison and the Wisconsin Alumni Research Foundation exist. To address this problem, the PatentsView team uses a disambiguation algorithm which aims to provide unique identifiers for assignees and inventors. A detailed explanation of the algorithm can be found in Monath, Jones, and Madhavan (2021). Although the algorithm solves the problem to a large extent, a small number of problematic assignee names still exist in the data. Therefore, the first intricacy involved in the data construction process is selecting university patents from the disambiguated assignee dataset. First, I determine the university names that will be chosen from the assignee dataset. For this, I use the list of university names in Hsu et al. (2021). There are 363 US universities on this list. Employing the direct utilization of university names may potentially result in the inadvertent exclusion of certain universities affected by the aforementioned issue. In response to this concern, I modify the list of university names to encompass a comprehensive representation of most universities. For example, I add ‘Stanford Jr’ and ‘Stanford Univ’ for Stanford University. Then, I select the assignees that contain the names in the modified list. Within the resultant dataset, the different expressions of a university name necessitated matching to a single name. To that end, I created another list of university names by using the names in the list mentioned above. The difference between these two lists is that the latter contains more general versions of the names, such as 'Wisconsin' for 'The Regents of the University of Wisconsin', 'University of Wisconsin-Milwaukee', or 'Board of Regents of the University of Wisconsin.' By iteratively traversing the dataset encompassing university names and the secondary list and verifying the presence of the names from the list within the dataset, I successfully matched the different versions of the university names to a single name.[3]

To calculate the citations of each patent received until the end of 2022, I use PatentsView's patent citation dataset. This dataset contains citations made to US patents by US patents.[4] Unfortunately, due to the unavailability of data, it is not feasible to ascertain the number of citations made to US patents originating from patents issued in other countries.[5] In the citation data, each row corresponds to a citation with citing patent's ID and cited patent's ID. As of December 2022, the dataset encompasses about 129 million citations. The university patents in our sample received 2,028,035 citations as of December 2022. Among 128,382 university patents, a subset of 80,179 patents, 62.4% of them, received at least one citation throughout their lifespan. The final sample encompasses patent and citation counts of 250 universities between 1975 and 2022. Figures 3.1 and Figure 3.2 depict the top 20 universities with the highest number of granted patents and average citations received.

3.1.1 Truncation Adjustment

The truncation of patents and citations presents challenges in analyzing innovative outputs and the quality of innovation. The truncation problem originates from the nature of patent and citation data. Typically, the count of granted patent applications (i.e., successful applications) is utilized for counts of patents in the economics of innovation literature. However, it generally takes about two years for a patent to be granted (Popp et al., 2003). Consequently, patents filed in recent years may not have been granted,

[3] I also tried some name-matching algorithms, such as fuzzy and Levenshtein distance. However, after manually checking the matched names, I concluded that these algorithms did not perform well.

[4] A patent is a US patent if it is granted by the USPTO. Both US citizens and foreigners can apply for a US patent.

[5] However, there is a dataset which provides citations made to foreign patents by granted US patents.

resulting in an underrepresentation of granted patent applications. Figure 3.3 depicts this truncation problem for the number of patents. Approximately 12.5-20% of those patents are still pending a grant decision as of December 2022.

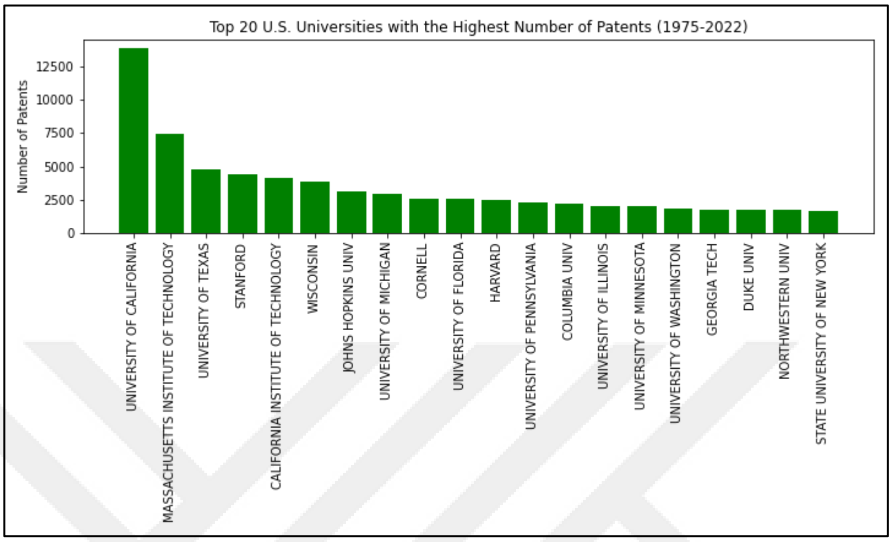

Figure 3.1: Top 20 Universities with the Highest Number of Granted Patent Applications

Notes: The "number of patents" refers to the cumulative count of successful (granted) patent applications within the period spanning from 1975 to 2022.

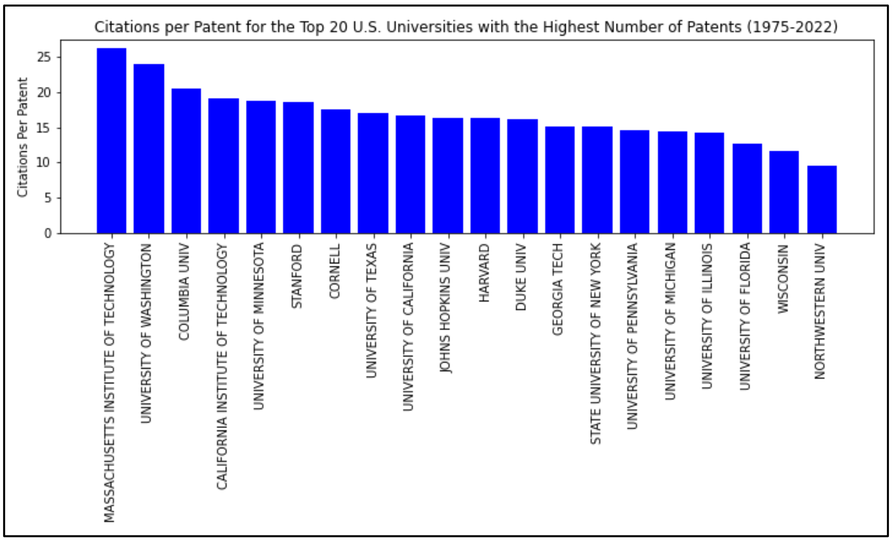

Figure 3.2: Citations per Patent

Notes: The citations per patent are computed by dividing the total number of citations received by the patents of the university by the total number of granted patent applications of that university.

Now, let us discuss the truncation of citations. As can be seen in Figure 3.4, the problem is even more severe for citations. Given that more recent patents have less time to receive citations compared to older patents, we face a truncation problem for citations. Consequently, employing the entire sample may cause biased conclusions. Moreover, any comparison between a recent and an older patent in terms of quality is ill-advised. Also, the number of patents and the number of citations vary across patent classes. To capitalize on this variation across patent classes, I use the International Patent Classification (IPC) system, which compromises eight major sections, namely human necessitates, performing operation operations, chemistry/metallurgy, textiles, fixed constructions, mechanical engineering (lighting, heating, weapons, blasting), physics, and electricity. Between 1970-2022, the physics and electricity classes have the highest number of patents, with the physics class accumulating 2,354,273 patents and the electricity class with 2,131,181 patents. Also, some classes are more extensively cited. For instance, patents in the human necessities class have 37.7 citations per patent, making it the most highly cited class between 1970-2022, followed by the fixed constructions class with 21.9 citations per patent. Thus, when comparing patents, one needs consider both year and class effects.

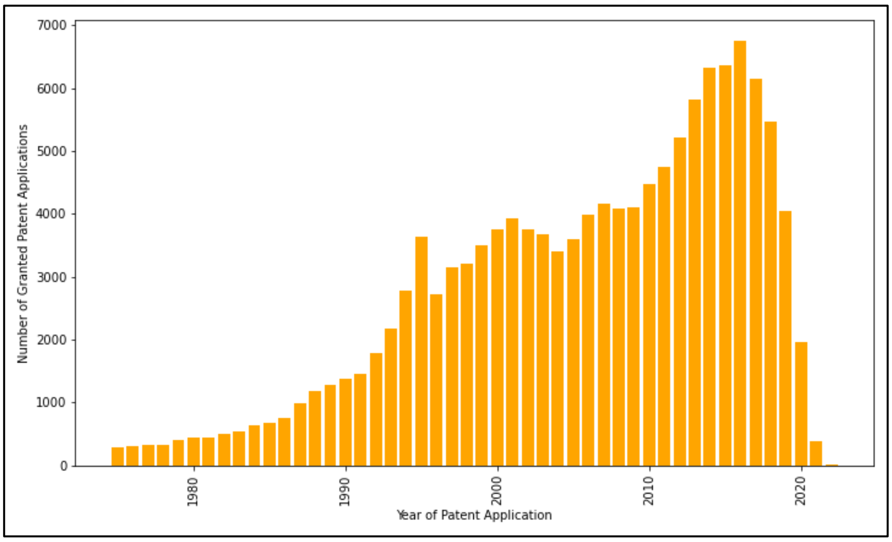

Figure 3.3: Successful (Granted) Patent Applications of US Universities by Application Year

As a remedy, we can limit our sample or use an adjustment method to address the truncation. I first restrict the sample until 2016. In the restricted sample, I only use the number of patents as the dependent variable, as this restriction is not enough to rectify the truncation issue for citations. Moreover, to leverage the full sample, I employ an adjustment method. In their seminal paper, Hall, Jaffe, and Trajtenberg (2001) present two adjustment methods for patent and citation bias. The first method, known as the "fixed effects" approach, is the one I adopt.[6] The fixed effects approach involves rescaling the number of patents and citations to take into account year-class effects. Adjusting the number of patents involves dividing the number of patents of a university in a patent class by the total number of patents in the same class. Hall et al. (2001) state that, "the rescaling purges the data of effects due to truncation, effects due to any systematic changes over time in the propensity to cite, and effects due to changes in the number of patents making citations" (p. 29). My methodology for calculating the time and class-adjusted number of patents and citations follows that of Lerner and Seru (2021).[7]

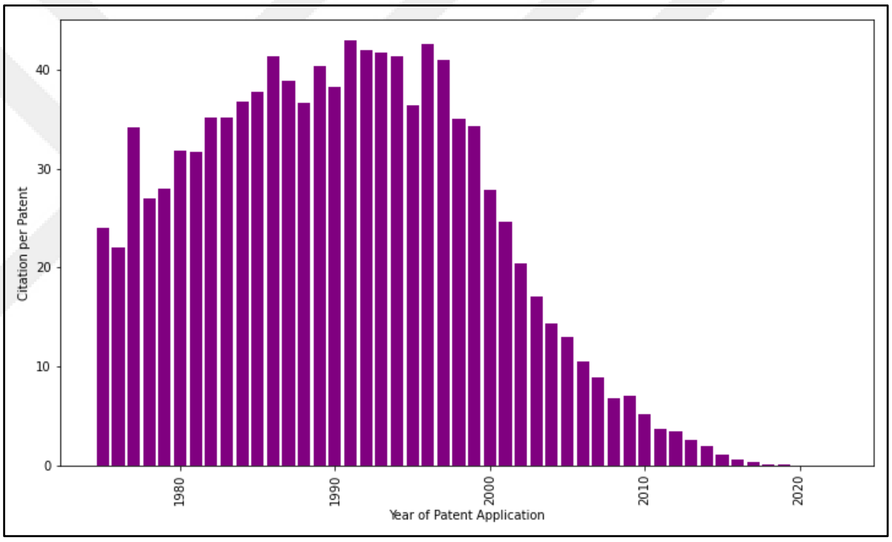

Figure 3.4: Citations per Patent for U.S. Universities by Application Year

The time and patent class-adjusted number of patents is computed as follows:

$$P_{it}^{*} = \sum_{k}^{K} \frac{p_{ikt}}{P_{kt}} \quad (1)$$

where P_{it}^{*} is the adjusted number of patents for university i applied for in year t, p_{ikt} is the total number of granted patents for university i applied for in year t in class k, P_{kt} is

[6] The other method is the "quasi-structural" approach which involves estimation of a functional form to differentiate the various effects on citations. This method makes two strong assumptions regarding the shape of the citation-lag distribution, namely proportionality and stationarity.

[7] Hall et al. (2001) discussed the fixed effect approach for citations whereas Lerner and Seru (2021) extended the discussion by including both the number of patents and citations.

the total number of granted patents applied for in year t in class k. K is the total number of IPC sections, which is eight and available for each patent in our dataset.

Similarly, adjusted number of citations is,

$$C_{it}^{*} = \sum_{k}^{K} \frac{\sum_{s}^{p_{ikt}} C_s}{\sum_{j}^{P_{kt}} C_j / P_{kt}} \tag{2}$$

where C_{it}^{*} is the adjusted number of citations received by university i for granted patents applied for in year t, C_s is the number of citations received by patent s of university i, belonging to class k. C_j is the number of citations received by patent j applied for in year t. As a case in point, if we would like to compare a 2010 patent, belonging to class k_1, with 8 citations to a 2019 patent, belonging to class k_2, with 2 citations, we divide them by the average number of citations received by patents in their class. Suppose the average number of citations of all granted patents applied for in 2010 in class k_1 is 10, while it is 1 for the patents applied for in 2019 in class k_2. Then, upon observing the adjusted number of citations, we can say that the 2019 patent is of better quality than the 2010 patent, although the 2010 patent received more citations.

As with any method, the "fixed effects" approach has weak points. Lerner and Seru (2021) discuss the inadequacy of this approach, and they propose the use of machine learning to predict the total citations that a patent will receive. Yet we do not adopt this strategy in this thesis.

3.2 *Student Enrollment Counts*

I create graduate enrollment counts in STEM fields, consisting of engineering, biological sciences/life sciences, mathematics, physical sciences, dentistry, and medicine associated with US universities using the Integrated Postsecondary Education System (IPEDS) data. The IPEDS dataset incorporates the counts of fall enrollments with information on the level of education (undergraduate, graduate, and first professional), residency classification (nonresident alien or domestic), and the major fields of study. The dataset spans 1980, 1986, and subsequent biennial intervals from 1994 to 2020. I consider both graduates and first professional degrees in the analysis.[8] While the graduate

[8] According to the US Department of Education International Affairs Office, a first-professional degree is an award that requires completion of a program that meets the following criteria: (1) completion

level consists of students in master's and doctoral programs, the latter refers mostly to students pursuing professional degrees in health-related fields. A 'unitid' number uniquely identifies each entity (university or campus) in the IPEDS data. There are 1,088 unique unitids in the dataset.

A discrepancy exists between the patent counts and student enrollment data in terms of the coverage of universities. In the patent data, universities with multiple campuses are consolidated and recorded as a single entity, while the student enrollment dataset incorporates each campus individually. For instance, the University of California, with ten distinct campuses, is recorded separately in the student enrollment dataset, whereas the patent counts dataset includes only a singular entry for ten campuses. Thus, before merging the two datasets, I combined different campuses of each university into a single entity. This consolidation required a manual examination to determine whether universities in the patent counts dataset have more than one unitid in the student enrollment dataset. 104 universities from the patent counts dataset have multiple campuses. After combining different campuses, I assigned a new unitid to universities with combined campuses. As the patent and student enrollment datasets solely share university names as a common column, I merged the datasets on university names. Some universities do not provide enrollment counts for specific years, resulting in missing values within the data. Therefore, I use the universities that provide enrollment counts for each year to ensure a more consistent structure for the analysis. The final student enrollment dataset contains 898 universities from 1986 and biennial intervals from 1994 to 2020. We use a subsample of this set by restricting the analysis to the universities with at least one granted patent in the sample period.

3.3 *Research and Development Spending*

I use the National Science Foundation's *Higher Education Research and Development Survey*, which contains annual research and development spending of US universities between 1972 and 2021. I utilize the R&D spendings in 1994 dollars for inflation adjustment.

of the academic requirements to begin practice in the profession; (2) at least 2 years of college work prior to entering the program; and (3) a total of at least 6 academic years of college work to complete the degree program, including prior required college work plus the length of the professional program itself. Some examples of first-professional degree include medicine, pharmacy, audiology, and dental science.

Our final dataset is a balanced panel that comprises counts of granted patent applications, citations received by the patents of universities, international and domestic graduate STEM enrollments, and research and development spending. The dataset encompasses 141 universities from 1986 and biennial intervals from 1994 to 2020. Table 3.1 presents the summary statistics of the main variables. An average university in the dataset generates 44.1 patents per year, with its patents receiving 282.4 citations per year. On average, the dataset shows an average enrollment of 472.5 foreign graduate STEM students whereas the graduate STEM enrollment for domestic students is 1188.9. Additionally, the universities in our dataset have an average R&D spending of $160,350,500.

Table 3.1: Summary Statistics

Variable	Obs	Mean	Std. Dev.	Min.	Max.
Patent Counts	2115	44.154	99.228	0	1291
Citation Counts	2115	282.364	1028.595	0	16499
Foreign Graduate STEM Enrollment	2115	472.509	718.614	0	8122
Domestic Graduate STEM Enrollment	2115	1188.871	1611.987	7	17331
Total Graduate Enrollment	2115	1661.38	2259.679	9	25289
R&D Spending	2115	160350.5	325934.1	0	4140518

Notes: R&D spendings are in thousands of dollars and in 1994 dollars. Patent and citation counts are sourced from the Patentsview data. Graduate STEM enrollment counts are obtained from the IPEDS data, while R&D spending information is collected from National Science Foundation's *Higher Education Research and Development Survey*.

Chapter 4:
METHODOLOGY

In this section, I present the main specifications and estimation techniques employed in the analysis. The study incorporates both Ordinary Least Squares (OLS) and Poisson methods for estimation. First, I describe the basic specifications. Subsequently, I outline the instrumental variables approach, which is employed to address the endogeneity problem in our specifications.

4.1 Basic Specifications

4.1.1 OLS Approach

To investigate the impact of international graduate STEM students on innovative outputs of universities, I first run the following specification:

$$innovation_{i,t} = \alpha_i + \psi_t + \gamma_i t + \beta_1 Foreign_{i,t-2} + \beta_2 Domestic_{i,t-2} + \beta_3 RD_{i,t-2} + \varepsilon_{i,t} \quad (3)$$

where $innovation_{i,t}$ represents the innovative output of university *i* in year *t*. I employ two measures of innovative output, one focusing on quantity, the other on quality of patents: the total patent count and the citations per patent received by university-affiliated patents. The count of patents for university *i* in year *t* represents the number of granted patent applications by university *i* in year *t*. To compute the citations per patent for patents of a particular university, one simply divides the total number of citations received by patents filed by university *i* in year *t* by the count of patent applications (granted) filed by that same university in the same year. Formally, the citations per patent received by university *i*'s patents filed in year *t* is

$$y_{it} = \frac{\sum_{j \in P_{it}} c_{ij}}{|P_{it}|} \quad (4)$$

where c_{ij} is the cumulative count of citations received by university *i* for patent *j*, as of December 2022. P_{it} denotes the set of granted patents filed by university *i* in year *t*, with $|P_{it}|$ representing the cardinality of set P_{it}. This metric evaluates a university's patent performance by encompassing both the quantity and quality of patents filed in a given year.

In the equation (3), α_i denotes university fixed effects while ψ_t represents year fixed effects. $\gamma_i t$ is university-specific time trends, $Foreign_{i,t-2}$ denotes the number of international students enrolled in graduate STEM programs in university *i* in year *t-2*, $Domestic_{i,t-2}$ is domestic graduate STEM enrollment in university *i* in year *t-2*, and $RD_{i,t-2}$ is R&D spending (in millions of dollars) of university *i* in year *t-2*.

As elaborated in Section 3.1.1, I initially use the unadjusted count of patents without accounting for truncation. In this scenario, our specification in (3) is applied to biennial intervals spanning from 1994 to 2016. Also, I incorporate the truncation adjusted counts of number of patents and citations, covering the biennial intervals from 1994 to 2020. In the subsequent models, I consistently designate 2016 as the final year of our sample. Nevertheless, in the results, I will present results encompassing both adjusted and unadjusted number of patents and citations.

Equation (3) aims to explain the variations in the quantity and quality of innovative outputs in a given university and year. This is achieved by exploiting foreign graduate STEM student enrollment two years before the given year while controlling for domestic graduate STEM enrollments, R&D spendings, university fixed-effects, year fixed-effects, and university-specific time trends. The university fixed-effects control time-invariant features of universities that influence their ability to produce patents, such as research culture, availability of research facilities, and location of the university. Certain universities focus on specific areas. For instance, Johns Hopkins is highly regarded for its medical and health-related research, while MIT is known for its excellence in science, engineering, and technology-related science. As patenting behavior differs across areas, specializing in certain areas affects the patent production of universities. Thus, controlling for these effects is necessary. Also, location of the university may have an important impact on patent production of that university. For example, Stanford University benefits from being situated in one of the world's leading innovation hubs, Silicon Valley. It is surrounded by major tech companies and startups, creating a favorable environment for research collaboration and patent production in high tech-fields. Thus, by incorporating university fixed-effects, I aim to disentangle the impact of foreign graduate STEM enrollments on innovative outputs from university-specific time-invariant characteristics influencing the quantity and quality of patents. On the other hand, year fixed-effects capture time-specific effects on patent production that are common across all universities. For instance, year fixed-effects can capture changes in government policies, technological advancement, or any other year-specific factors that could influence patent production. The university-specific time trends capture any linear changes in the time-varying university-specific factors that impact innovative outputs. The relationship between R&D investments and patent production has long been studied, such as Pakes and Griliches

(1980). To account for this relationship, I incorporate the universities' R&D spendings as a control variable in the main specification.

The impact of graduate students on production of innovative outputs is not immediate, as the patent production process involves the lag between the start of research and the patent application. Additionally, there is a further delay between the patent application and the granting of the patent. However, I use the application date of the patents to determine the time of the innovation, so it is unnecessary to account for the latter lag. This approach aligns with several notable studies, such as those conducted by Kerr and Lincoln (2010), Hunt and Gauthier-Loiselle (2010), and Chellaraj et al. (2008), which also utilize patent application years as indicators of innovation timing. The lag between the start of research and the patent application is approximately 18 months (Stephan and Levin, 1992). Considering the time required for students to become actively engaged in the research process, it is reasonable to assume a lag of 2 years. Moreover, the impact of R&D spending on innovation production may manifest with a lag. For example, Gurmu et al. (2010) propose lags of 4 years whereas Stuen et al. (2012) suggest a lag of 1 year for R&D spending. Drawing from this literature, I use lags of 2 years for R&D spendings in the analysis.

To account for the biennial structure of the student enrollment data, I adopt a simple approach in calculating the number of patents, number of citations, and R&D spendings. Starting from 1993 until 2020, I add odd years' values to subsequent even years' values to obtain comprehensive numbers because using only the number of patents and citations in even years may underestimate the impact of international students. For instance, I add counts of patents in 1993 to the number of patents in 1994; counts of patents in 1995 to 1996, and so forth. The rationale behind this approach is that international students enrolled in year *t-2* potentially affect patent production in both year *t-1* and *t*. Subsequently, through the application of this approach, we can encapsulate the impact of enrollment in *t-2* on the combined count of patents in *t-1* and *t*, where *t* is an even year. The same reasoning applies to the number of citations and R&D spendings.

4.1.2 Fixed- Effects Poisson Approach

Since one of our dependent variables, the count of patent applications (granted), is a nonnegative count variable, especially with a substantial frequency of zeros, it is attractive to use a fixed effects (FE) Poisson estimator (Wooldridge, 2019). Figures 5 shows the histogram of number of patents which reveal the prevalence of zero counts in our dataset.

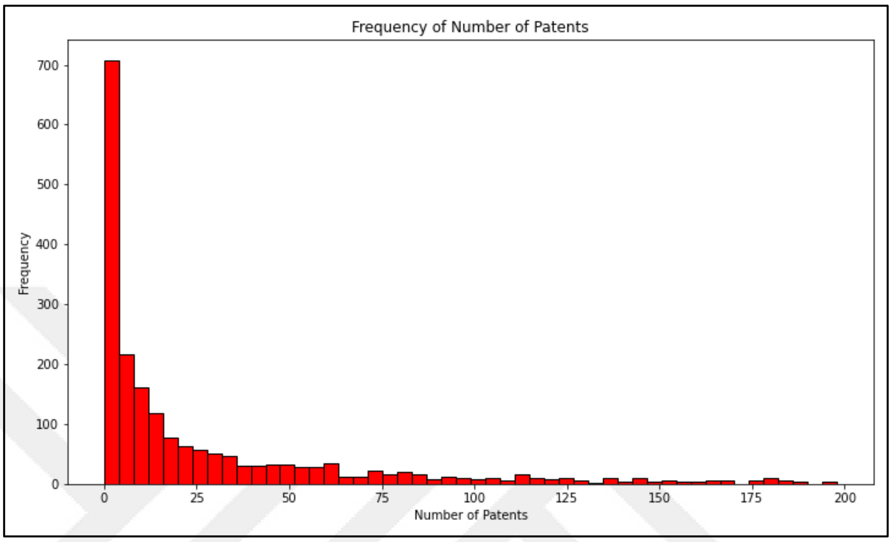

Figure 4.1: Histogram of Number of Patents

The fixed effects Poisson estimator was introduced by Hausman, Hall, and Griliches (1984) for analyzing count panel data models with fixed effects. As elaborated by Wooldridge (1999) and Wooldridge (2010, Chapter 19), the FE Poisson is an appealing estimator due to its consistency and asymptotic normality properties. Overdispersion could be a concern in our model since the variance of patent counts is much higher than its mean, as illustrated in Table 1. Nonetheless, Wooldridge (1999) demonstrates that the FE Poisson estimator is completely robust to every failure of the Poisson assumptions, including the instances where the mean and variance are not equal. The only requirement for this estimator's validity is the correct specification of the conditional mean. Consequently, there is no need for us to be concerned about the overdispersion issue.

An alternative approach for count panel data is the fixed effects negative binomial (FENB) model. However, the FENB model has certain limitations. As highlighted by Guimãres (2008), the FENB model does not control for individual fixed effects unless a very specific set of assumptions is satisfied. Moreover, Wooldridge (1999) cautions that the FENB model is notably influenced by the specification of conditional expectation or conditional variance of the dependent variable. Given these limitations of the FENB model and considering the merits and drawbacks of both approaches, I choose the FE Poisson over the FENB model.

The first step of the FE Poisson approach is to specify the conditional mean for the dependent variable:

$$\mu_{i,t} = E\left[innovation_{i,t} \middle| \boldsymbol{X}_{i,t}, c_i\right] = c_i exp\left(\beta_1 Foreign_{i,t-2} + \beta_2 Domestic_{i,t-2} + \beta_3 RD_{i,t-2} + \psi_t + \gamma_i t\right) \quad (5)$$

where $innovation_{i,t} \sim Poisson\left(\mu_{i,t}\right)$ and $\boldsymbol{X}_{i,t}$ is a matrix that contains $Foreign_{i,t-2}$ and other control variables in (1), c_i is the time-invariant unobserved heterogeneity. Hausman et al. (1984) use the conditional maximum likelihood method for estimation. Let $n_i \equiv \sum_{t=1}^{T} innovation_{i,t}$. They show that

$$\boldsymbol{innovation}_i \sim Multinomial\left(n_i, p_1(\boldsymbol{X}_i, \boldsymbol{\beta}), \dots, p_T(\boldsymbol{X}_i, \boldsymbol{\beta})\right) \quad (6)$$

where $\boldsymbol{innovation}_i = \left(innovation_{i,1}, innovation_{i,2}, \dots, innovation_{i,T}\right)'$ is a Tx1 vector, and

$$p_t(\boldsymbol{X}_i, \boldsymbol{\beta}) \equiv \frac{exp\left(\boldsymbol{X}_{i,t}\boldsymbol{\beta}\right)}{\sum_{j=1}^{T} exp\left(\boldsymbol{X}_{i,j}\boldsymbol{\beta}\right)} \quad (7)$$

In our case, $t = 1994, 1996, \dots, 2016$. Then, we estimate $\boldsymbol{\beta}$ by employing the conditional maximum likelihood. The conditional log likelihood for observation *i* is

$$l_i(\boldsymbol{\beta}) = \sum_{t=1}^{T} innovation_{i,t} log[p_t(\boldsymbol{X_i}, \boldsymbol{\beta})] \quad (8)$$

Finally, the fixed effects Poisson estimator is $\widehat{\boldsymbol{\beta}}$ that maximizes $\sum_{i=1}^{N} l_i(\boldsymbol{\beta})$, where N is the number of universities in our sample.

4.2 *Instrumental Variables Approach*

In our basic specification, concerns regarding both reverse causality and omitted variables might arise, i.e., $E[\varepsilon|foreign] \neq 0$. For example, a rise in the number of patents or citations of a university could attract more international graduate students. This correlation might originate from the notion that an increase in the quantity or quality of patents usually reflects a university's strong research competence and reputation. This relationship could introduce the problem of reverse causality, where the causality is bidirectional. Furthermore, it is important to consider the unobservable characteristics of universities, as pointed out by Stuen et al. (2012), vary non-linearly over time. These characteristics potentially impact both the patent production and international graduate student enrollments. For instance, an increase in faculty quality could allure more international graduate students and increase the number of patents or citations.

My approach to address the endogeneity problem is to employ a shift-share (or "Bartik") instrument.[9] This entails combining a university's predetermined share of nationwide international graduate STEM enrollment with the national level of supply of international graduate STEM students. The instrument for university *i*'s international graduate STEM enrollments in year *t* is

$$B_{i,t} = \frac{S_{i,1986}}{S_{1986}} S_t \tag{9}$$

Here, $S_{i,1986}$ represents the international graduate STEM enrollment at university *i* in 1986, S_{1986} denotes the total international graduate STEM enrollment in the US in 1986. Thus, the "share", $\frac{S_{i,1986}}{S_{1986}}$, signifies the proportion of university *i* in 1986's aggregate international graduate STEM student enrollments across all US universities. S_t (the "shift") is the aggregate flow of international graduate STEM students to the US in year *t*. $B_{i,t}$ is therefore the predicted value of the international graduate STEM enrollment for a university *i* in year *t*, assuming the distribution of international graduate students among

[9] Shift-share (or "Bartik") instruments were developed by Bartik (1991) where he interacted local industry shares and national industry growth rates to estimate the impact of labor demand shifts on wages. Some recent prominent applications of shift-share instruments include Autor et al. (2013), Card (2009), Hunt and Gauthier-Loiselle (2010), Jaravel (2019), and Oberfield and Raval (2021).

universities aligns with the distribution observed in 1986. Goldsmith-Pinkham et al. (2020) find that shift-share instruments are equivalent to employing GMM estimation with the shares serving as instruments. As a result, the primary source of identification originates from the shares rather than the shifts. The share has predictive values of incoming international graduate student flows, as certain universities have historically attracted larger numbers of international students, while others have attracted less. Table 2 demonstrates that the international graduate STEM enrollments in 1986 is highly predictive of enrollments in recent years. Not only do we observe statistical significance, but also notable R^2 values ranging from 0.72 to 0.75. These values indicate that 72% to 75% of variation in the international student enrollments in recent years can be explained by the international student enrollment in 1986. Figure 4.2 depicts the top 20 universities with the highest average yearly international graduate STEM student enrollments between 1986-2020. Also, as predetermined shares remain unaffected by current factors that could impact the immigration location choices of students, i.e., universities, their incorporation alongside nationwide trends in international graduate STEM student inflows helps to disentangle international student inflows from the fluctuations in university-specific characteristics.

Table 4.1:OLS Regressions of International Graduate STEM Enrollments in Recent Years on Enrollments in 1986

	(1)	(2)	(3)	(4)
Int. Grad. STEM Enrollments	2020	2018	2016	2014
Enrollment in 1986	2.419***	2.650***	2.604***	2.312***
	(0.126)	(0.132)	(0.130)	(0.120)
Constant	21.87	32.77	59.35	88.80*
	(52.35)	(54.74)	(53.73)	(49.65)
Observations	141	141	141	141
R^2	0.725	0.743	0.744	0.728

Notes. *Significant at 10%, **Significant at 5%, ***Significant at 1%.

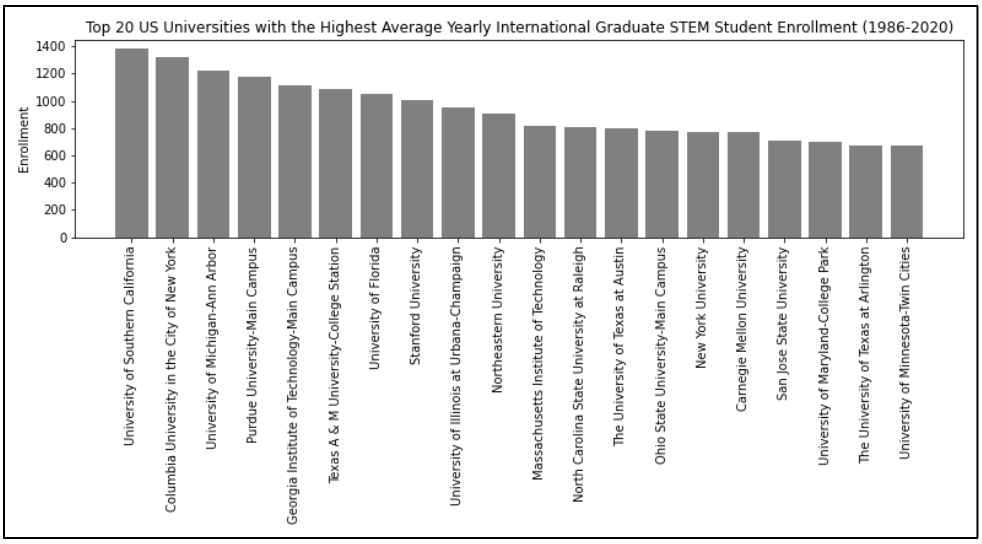

Figure 4.2: Top 20 Universities with the Highest Average Yearly International Graduate STEM Enrollment

Note**:** Average yearly student enrollments are computed using the IPEDS data collected at biennial intervals spanning from 1986 to 2020.

I also investigated an alternative instrument, utilizing the fluctuations in enrollments across STEM majors (engineering, biological sciences/life sciences, mathematics, physical sciences, dentistry, and medicine). This approach yielded insignificant results (see the Results section). This alternative instrument is defined as follows:

$$\tilde{B}_{i,t} = \sum_{k} \frac{S_{i,k,1986}}{S_{k,1986}} S_{k,t} \tag{10}$$

where $\frac{S_{i,k,1986}}{S_{k,1986}}$ is university i's share in 1986 of the nationwide international graduate student enrollments in major k, and $S_{k,t}$ is the nationwide international graduate student enrollment in major k in year t. This instrument captures the variation in international graduate STEM enrollments stemming from divergent effects of nationwide international graduate STEM enrollments on universities, influenced by the universities' unique enrollment trends across various majors. For instance, if a particular university has a historical tendency to recruit a higher number of international graduate students in certain majors, then a shock to the international graduate STEM student flows in this major would exert a relatively higher influence on that specific university's innovative outputs.

I employ two distinct estimation methods: Two-Stage Least Squares (2SLS) and a control function approach for FE Poisson. The 2SLS approach is widely used and does not require a detailed discussion. However, I find it imperative to delve into the control function approach as its application is less common.

4.2.1 A Fixed-Effects Poisson/Control Function Approach for Count Data Models with Endogenous Regressors

Control function (CF) approach is employed for correcting endogeneity in linear and nonlinear unobserved effects models. A CF is defined by Barnow et al. (1981) as "a variable that, when added to a regression, renders a policy variable appropriately exogenous." As Wooldridge (2015) states, "the CF approach is inherently an instrumental variables method." Thus, we need one or more instrumental variables to construct a CF. In this context, I use residuals from the first-stage regression as a CF. Cameron and Trivedi (2009, Chapter 17) and Wooldridge (2015) discuss the applications of the CF approach to address endogeneity in count data models with cross-sectional data framework. Meanwhile, Lin and Wooldridge (2019) extend this discussion to linear and nonlinear models, particularly focusing on fixed effects models. For this analysis, I adopt the approach outlined by Lin and Wooldridge. This method combines FE Poisson with the CF approach.

Different than the mean specification in (2), now we have a time-varying omitted factors that we suspect are correlated with our main regressor, international graduate STEM enrollment. This leads to the following model,

$$E\left[innovation_{i,t} \middle| Foreign_{i,t-2}, \boldsymbol{x}_{i,t-2}, c_i, r_{i,t}\right] = c_i exp\left(\beta_1 Foreign_{i,t-2} + \boldsymbol{x}_{i,t-2}\boldsymbol{\beta_2} + r_{i,t}\right) \quad (11)$$

In this equation, $innovation_{i,t}$ represents the count of granted patent applications of university *i* in year *t*, $Foreign_{i,t}$ stands for the international graduate STEM enrollment in university *i* in year *t*. $\boldsymbol{x}_{i,t-2}$ is a row vector that encompasses the exogenous variables: 2-year lagged university-specific time trends, year fixed effects, RD spending, and domestic enrollment. $r_{i,t}$ denotes the time-varying omitted factors that we suspect to be correlated with $Foreign_{i,t}$. I denote the excluded instrument as $B_{i,t}$, which is the predicted foreign enrollment as constructed based on shift-share method in our analysis.

I specify a reduced form for $Foreign_{i,t}$,

$$Foreign_{i,t} = \boldsymbol{z}_{i,t}\boldsymbol{\Psi}_1 + f_i + u_{i,t} \tag{12}$$

where $\boldsymbol{z}_{i,t} = \left(\boldsymbol{x}_{i,t}, B_{i,t}\right)$ and f_i is unobserved heterogeneity. Assume that $\{\boldsymbol{z}_{i,t}\}$ is strictly exogenous, so the endogeneity of $Foreign_{i,t}$ stems from the correlation between $\{u_{i,t}\}$ and $\{r_{i,t}\}$. I proceed by assuming $\{u_{i,t}\}$ and $\{r_{i,t}\}$ are related via:

$$r_{i,t} = \rho u_{i,t} + \omega_{i,t} \tag{13}$$

where $\boldsymbol{\omega}_i = (\omega_{i,1}, \dots, \omega_{i,T})'$ and is independent of $\boldsymbol{u}_i = (u_{i,1}, \dots, u_{i,T})'$. I also assume that $\boldsymbol{r}_i = (r_{i,1}, \dots, r_{i,T})'$ and $\boldsymbol{u}_i$ are independent of (c_i, f_i) where $\boldsymbol{r}_i$ is the vector of omitted variables in (11) and $\boldsymbol{u}_i$ is the vector of residuals from (12). For example, we can interpret $u_{i,t}$ as a measure of the faculty quality at university *i* in year *t*, a factor that could influence the enrollment of international students at that university. On the other hand, $r_{i,t}$ can represent the research collaboration undertaken by the university with other academic institutions or firms. Given that faculty members with higher quality often have more extensive networks in both academia and the private sector, their existence in the university can potentially increase the research collaboration with other institutions. Consequently, the increased research collaboration could influence the university's patent production. Therefore, the correlation between $Foreign_{i,t}$ and $r_{i,t}$ is fundamentally a result of the correlation between $r_{i,t}$ and $u_{i,t}$. Substituting (13) in (11) yields:

$$E\left[innovation_{i,t}\middle|Foreign_{i,t-2}, \boldsymbol{x}_{i,t-2}, c_i, \boldsymbol{u}_i\right] = c_i exp\left(\beta_1 Foreign_{i,t-2} + \boldsymbol{x}_{i,t-2}\boldsymbol{\beta}_2 + \rho u_{i,t} + \omega_{i,t}\right) \tag{14}$$

In (14), we can view $u_{i,t}$ as an explanatory variable. By including $u_{i,t}$ to the model, we obtain a new error term, $\omega_{i,t}$. This new error term is uncorrelated with $Foreign_{i,t-2}$. By virtue of our assumption that $\boldsymbol{r}_i$ and $\boldsymbol{u}_i$ are uncorrelated with $\boldsymbol{z}_i = (\boldsymbol{x}_i, B_i)$, it follows that $\boldsymbol{\omega}_i$ is also uncorrelated with $\boldsymbol{z}_i$. Consequently $\boldsymbol{\omega}_i$ must exhibit no correlation with $Foreign_{i,t}$ for all time periods *t*. Therefore, including $u_{i,t}$ in the equation "controls for" the endogeneity of $Foreign_{i,t-2}$. Given that $u_{i,t}$ is unobservable, it is essential to replace it with a consistent estimate.

Following the two-step estimation:

1. Estimate the equation (12) by fixed effects and obtain the fixed effects residuals,

$$\hat{\ddot{u}}_{i,t} = \ddot{Foreign}_{i,t} - \ddot{\boldsymbol{z}}_{i,t}\widehat{\boldsymbol{\Psi}}_1$$

where $\ddot{Foreign}_{i,t} = Foreign_{i,t} - \frac{1}{T}\sum_{j=1}^{T} Foreign_{i,j}$.

2. Use FE Poisson on

$$E\left[innovation_{i,t}\middle|Foreign_{i,t-2}, \boldsymbol{x}_{i,t-2}, \hat{\ddot{u}}_{i,t-2}, c_i\right]$$
$$= c_i \exp(\beta_1 Foreign_{i,t-2} + \boldsymbol{x}_{i,t-2}\boldsymbol{\beta}_2 + \rho\hat{\ddot{u}}_{i,t-2})$$

to estimate β_1 consistently.

Chapter 5:
RESULTS

In this section, I present the outcomes of regressions that delve into the impact of international graduate STEM students on patent-related performances of universities, encompassing both the patent quantity and quality measures, namely the number of patents and citations per patent. The initial analysis focuses on the patent counts and citation counts, spanning the biennial periods from 1994 to 2016. Subsequently, I present the findings related to the truncation adjusted dependent variables, which cover the biennial intervals from 1994 to 2020, in the robustness checks section. Finally, additional robustness checks are presented in section 5.3.2.

5.1 *The Impact of International Graduate STEM Students on Patenting*

Table 5.1 presents a comprehensive array of estimation techniques: Ordinary Least Squares (OLS), Two-Stage Least Squares (2SLS), Fixed Effects (FE) Poisson, and FE Poisson with Control Function (CF). Columns 1 and 2 of Table 5.1 report the OLS and 2SLS estimates of the benchmark model in (3). Columns 3 and 4 show the FE Poisson and Poisson with CF estimates of the equations (5) and (11), respectively. These estimates show the impact of international graduate STEM students on patenting associated with each university for the period 1994-2016. I quantify patenting based on the number of granted patent applications. Our variable of interest is 2-year lagged counts of international graduate STEM enrollment in the university. Each model controls for domestic graduate STEM enrollments, R&D spending, university fixed effects, year fixed effects, and university-specific linear trends. The patent quantity and quality of a university tend to be correlated throughout the period I analyze. Thus, I cluster the standard errors at the university level to allow for correlation across periods within each university.

In the OLS regression analysis (reported in column 1), the effect of international graduate STEM students on patent production of universities stands as statistically insignificant. Furthermore, the impact of domestic graduate STEM students and R&D spending on patenting are not different than zero, statistically. In column 2 where I present the 2SLS estimates, a contrast becomes evident. In contrast to the OLS scenario, the 2SLS method reveals a statistically significant impact of international students on the patent counts. On average, each additional international graduate STEM student adds 0.05 patents per year, translating to a cumulative increase of 0.3 patents over the duration of a six-year doctoral program. The power of the instrumental variable in the 2SLS model is highlighted by an F-statistics of 17.56, which increases confidence against the concerns of a weak instrument problem.

OLS, 2SLS, FE Poisson, and FE Poisson with CF Approach Estimates of the Impact of International Graduate STEM Students on Patenting

	(1)	(2)	(3)	(4)
Dependent Variable:	Number of Patents	Number of Patents	Number of Patents	Number of Patents
Method:	OLS	2SLS	FE Poisson	FE Poisson with CF Approach
2-Yr. Lagged Int. Graduate STEM Students	0.0111	0.0557***	0.000075	0.000112**
	(0.0116)	(0.0166)	(0.000049)	(0.000050)
2-Yr. Lagged Dom. Graduate STEM Students	0.00220	-0.00473	-0.000014	-0.000020
	(0.00663)	(0.00952)	(0.000017)	(0.000022)
2-Yr. Lagged R&D Spending	-0.0317	-0.0349	-0.000117	-0.000115
	(0.0581)	(0.0589)	(0.000095)	(0.000095)
Observations	1,551	1,551	1,551	1,551
R^2	0.638	0.619		
Number of Universities	141	141	141	141
University FE	YES	YES	YES	YES
Year FE	YES	YES	YES	YES
University-Spec. Trends	YES	YES	YES	YES
F-Stat of IV		17.56		

Notes: *Significant at 10%, **Significant at 5%, ***Significant at 1%. Robust standard errors in parentheses and clustered by university. Dependent variable, $innovation_{i,t}$, is the number of granted patents applied for by university *i* in year *t*. All models control for domestic graduate STEM enrollments, R&D spending, university fixed effects, year fixed effects, and university-specific linear trends. The sample contains 141 universities and spans the biennial intervals from 1994 to 2016. The number of patents considered in the analysis obtained from the *PatentsView* database. Student enrollment counts are sourced from the *IPEDS* database, while R&D spending are computed from National Science Foundation's *Higher Education Research and Development Survey*. R&D spending are in millions of dollars and in 1994 dollars. To reflect the biennial structure of our sample, the counts of patents and R&D spending are derived by combining values from odd years with those from even years, from 1993 to 2016.

Comparing the OLS and 2SLS models underscores a notable discrepancy: the OLS approach tends to underestimate the impact of international students on patent production.

In line with these findings, Stuen et al. (2012) similarly observe that the OLS approach underestimates the impact of both international and domestic graduate STEM students on a university's publication count. They point out that this underestimation implies that when other pivotal factors shape patenting, such as availability of research funding, were relatively low, universities tended to receive higher volumes of student applications. This observation underscores the complex interplay of university-related determinants of patenting and their implications for attracting students. Kahn (2010) supports this finding by demonstrating that applications to graduate schools increase during recessions.

As discussed in section 3, the nature of the count data within our sample supports the application of a Poisson regression model. Initially, employing the FE Poisson estimator, similarly to the OLS scenario, produces statistically insignificant coefficient estimators. However, when I employ FE Poisson with CF to address the endogeneity issue, the estimator achieves statistical significance at 5%. The FE Poisson model in (5) implies that

$$\frac{\partial E\left[innovation_{i,t} \middle| \boldsymbol{X}_{i,t}, c_i\right]}{\partial foreign_{i,t-2}} = \beta_1 \times E\left[innovation_{i,t} \middle| \boldsymbol{X}_{i,t}, c_i\right] \tag{15}$$

Therefore, β_1 can be interpreted as the semi-elasticity of $E\left[innovation_{i,t} \middle| \boldsymbol{X}_{i,t}, c_i\right]$ with respect to $foreign_{i,t-2}$. In column (4) of Table 5.1, the coefficient of $foreign_{i,t-2}$ is 0.00011 which implies that an additional international graduate STEM student is associated with a 0.00011 proportionate increase, or a 0.011% rise in the number of patents.

Although the extent of the impact of international graduate STEM students on patenting is small, these estimations suggest that international graduate STEM students do have a positive influence on university-level U.S. innovation. This aligns with the findings of Stuen et al. (2012), who similarly detect a minor yet positive effect of international students on patenting activity within an OLS regression context. However, unlike our results, their IV method does not yield statistically significant estimators.

5.2 *The Impact of International Graduate STEM Students on Patent Quality*

Table 5.2 presents the estimates of the impact of international students on patent quality, with the dependent variable being citations per patent. These estimations shed light on the impact of international graduate STEM students on the quality of innovation at each university spanning the biennial intervals from 1994 to 2016. The assessment of innovation quality at a specific university during a particular year is quantified based on the citations per patent, as defined in equation (4). The variable of interest is 2-year lagged counts of international graduate STEM enrollment in the university. Each model controls for domestic graduate STEM enrollments, R&D spending, university fixed effects, year fixed effects, and university-specific linear trends. Standard errors are clustered at the university-level.

These findings suggest that international graduate STEM student do not exert a statistically significant influence on patent quality. In this analysis, I evaluate the effect of international graduate STEM students on the cumulative count of citations received by patents filed two years after the enrollment of international graduate STEM students. In sum, based on the current analysis, it can be concluded that international graduate STEM students positively impact the patent counts, but this impact does not extend to the quality of patents.

Table 5.1: OLS and 2SLS Estimates of the Impact of International Graduate STEM Students on Patent Quality

	(1)	(2)	(3)	(4)
Dependent Variable:	Citations per Patent	Citations per Patent	Citations per Patent	Citations per Patent
Method:	OLS	2SLS	FE Poisson	FE Poisson with CF Approach
2-Yr. Lagged Int.	-0.0013	0.0175	0.000011	0.000928
Graduate STEM Students	(0.0028)	(0.0109)	(0.000275)	(0.000741)
2-Yr. Lagged Dom.	0.0005	-0.0025	-0.000011	-0.000157
Graduate STEM Students	(0.0011)	(0.0020)	(0.000112)	(0.000170)
2-Yr. Lagged R&D	0.0053	0.0040	0.000206	0.000002
Spending	(0.0036)	(0.0038)	(0.000284)	(0.000346)
Observations	1,551	1,551	1,485	1,485
R^2	0.4961	0.4825		
Number of Universities	141	141	135	135
University FE	YES	YES	YES	YES
Year FE	YES	YES	YES	YES
University-Spec. Trends	YES	YES	YES	YES
F-Stat of IV		17.56		

Notes: *Significant at 10%, **Significant at 5%, ***Significant at 1%. Robust standard errors in parentheses and clustered by university. Dependent variable is the citations per patent received by university i's patents filed in year t. This variable is computed by taking the cumulative count of citations received by university i's granted patents filed in year t, as of December 2022, and dividing it by the total number of granted patent applications of that university in year t. All models control for domestic graduate STEM enrollments, R&D spending, university fixed effects, year fixed effects, and university-specific linear trends. The sample contains 141 universities and spans the biennial intervals from 1994 to 2016. The citation counts obtained from the *PatentsView* database. To reflect the biennial structure of our sample, the counts of citations and R&D spending are derived by combining values from odd years with those from even years, from 1993 to 2016.

5.3 *Robustness Checks*

5.3.1 Results with Truncation Adjusted Dependent Variables

The truncation of patent counts and citations present challenges in analyzing the innovative outputs, as elaborated in section 3.1.1. Up to this point, I have employed a restricted sample spanning from 1994 to 2016, although the full sample extends to 2020. To benefit from the full sample, I implement truncation adjustment methods for the count of patents and citations, as defined in equations (1) and (2), respectively. Now, I present the results with truncation adjusted number of patents and citations as displayed in Table 5.3.

The adjustment method changes the scale of dependent variables. For instance, the mean of the adjusted number of patents is 0.0014. Thus, the interpretation of the coefficient estimates may not serve our goals. Similar to the findings in Table 5.1, the OLS method produces insignificant estimates of the impact of international students on the number of patents. In contrast, the 2SLS method yields a positive and significant coefficient estimate. However, Poisson regressions do not yield significant estimates in this scenario. Regarding the adjusted citations per patent, the results are similar to those in Table 5.2. Specifically, it is observed that international graduate STEM students do not exert a significant impact on patent quality.

Table 5.2: Estimates of the Impact of International Graduate STEM Students on Patent Production and Quality with Truncation Adjusted Dependent Variables

Panel A: Patents	(1) Adjusted Number of Patents	(2) Adjusted Number of Patents	(3) Adjusted Number of Patents	(4) Adjusted Number of Patents
Method:	OLS	2SLS	FE Poisson	FE Poisson with CF
2-Yr. Lagged Int. Graduate STEM Students	0.0001	0.0024**	0.0639	0.0937
	(0.0002)	(0.0009)	(0.0409)	(0.0984)
2-Yr. Lagged Dom. Graduate STEM Students	-0.0001	-0.0004**	-0.0299	-0.0349
	(0.0002)	(0.0002)	(0.0249)	(0.0387)
2-Yr. Lagged R&D Spending	-0.0005	-0.0005	-0.2082*	-0.2072*
	(0.0009)	(0.0007)	(0.1153)	(0.1171)
Observations	1,833	1,833	1,794	1,794
R^2	0.4295	0.2741		
Number of Universities	141	141	138	138
University FE	YES	YES	YES	YES
Year FE	YES	YES	YES	YES
University-Spec. Trends	YES	YES	YES	YES
F-Stat of IV		17.56		
Panel B: Citations	(1) Adjusted Citations per Patent	(2) Adjusted Citations per Patent	(3) Adjusted Citations per Patent	(4) Adjusted Citations per Patent
Method:	OLS	2SLS	FE Poisson	FE Poisson with CF
2-Yr. Lagged Int. Graduate STEM Students	0.0307	0.2634	0.1352	1.1041
	(0.0895)	(0.2873)	(0.2711)	(0.7202)
2-Yr. Lagged Dom. Graduate STEM Students	0.0271	-0.0088	0.0317	-0.1208
	(0.0766)	(0.0899)	(0.1304)	(0.1764)
2-Yr. Lagged R&D Spending	0.0560	0.0604	0.4169	0.3697
	(0.1382)	(0.1417)	(0.2767)	(0.2689)
Observations	1,833	1,833	1,755	1,755
R^2	0.1196	0.1183		
Number of Universities	141	141	135	135
University FE	YES	YES	YES	YES
Year FE	YES	YES	YES	YES
University-Spec. Trends	YES	YES	YES	YES
F-Stat of IV		17.56		

Notes. *Significant at 10%, **Significant at 5%, ***Significant at 1%. Robust standard errors in parentheses and clustered by university. Dependent variables are adjusted for truncation using the methods described in section 2.1.1. All models control for domestic graduate STEM enrollments, R&D spending, university fixed effects, year fixed effects, and university-specific linear trends.

5.3.2 *Other Robustness Checks*

Table 5.4 shows the results of various robustness checks assessing the impact of international graduate STEM students on number of patents. These checks include the utilization of the alternative instrument, denoted as $\tilde{B}_{i,t}$ in equation (10), the use of a binary dependent variable indicating whether the university has a patent in a given year, and the use of the share of international graduate STEM students in the university as a substitute for the counts of international and domestic graduate STEM student enrollments.

In columns 1 and 2 of Table 5.4, I instrument international graduate STEM student enrollments with the alternative instrumental variable, $\tilde{B}_{i,t}$. This instrument captures the variation in international graduate STEM enrollments arising from diverse effects of nationwide international graduate STEM enrollments on universities. These impacts are shaped by the distinct enrollment trends that universities experience across various majors. The main instrument, $B_{i,t}$, does not capture the variations in these majors. Columns 3 and 4 present the results of regressions where the share of international students in the total number of graduate STEM students serves as the main regressor. In the last column, I present the outcomes of regressions using a binary dependent variable labeled “Patented”. This variable takes the value 0 if the university does not have any patents and 1 if the university holds at least 1 patent in a given year. The inclusion of this binary dependent variable aims to assess whether international graduate STEM students influence the patenting behavior of universities, independent of their impact on the magnitude of patent production.

The statistically significant and positive findings observed in Table 5.1 regarding the impact of international graduate students on patenting are not robust in these alternative specifications in the columns 1 through 4. On the other hand, binary outcome IV results provide suggestive evidence that international graduate STEM students affect the extensive margin of patenting. I do not examine the robustness of the impact of international students on patent quality, as my main regression models yield insignificant estimates for this impact.

Table 5.3: Robustness Checks

	Alternative IV	Alternative IV	Int. Graduate Share	Int. Graduate Share	Binary Outcome
Dependent Variable:	(1) Number of Patents	(2) Number of Patents	(3) Number of Patents	(4) Number of Patents	(5) Patented
Method:	2SLS	FE Poisson with CF	2SLS	FE Poisson with CF	2SLS
Int. Graduate STEM Students	-0.01694	-0.00003			0.10471
	(0.04603)	(0.00007)			(0.14312)
Domestic Graduate STEM Students	0.00655	0.000005			-0.01319
	(0.01371)	(0.00003)			(0.02359)
R&D Spending	-0.02973	-0.00012	-0.02679	-0.00012	-0.00156
	(0.05455)	(0.00009)	(0.05547)	(0.00017)	(0.02015)
Int. Graduate STEM Student Share			0.28666	-0.81734	
			(3.16344)	(18.73193)	
Observations	1,551	1,551	1,551	1,551	1,551
R^2	0.63038		0.43246		0.22322
University FE	YES	YES	YES	YES	YES
Year FE	YES	YES	YES	YES	YES
University-Spec. Trends	YES	YES	YES	YES	YES
Number of Universities	141	141	141	141	141
F-Stat of IV	17.06		17.56		17.56

Notes: *Significant at 10%, **Significant at 5%, ***Significant at 1%. Robust standard errors in parentheses and clustered by university. All models control for domestic graduate STEM enrollments, R&D spending, university fixed effects, year fixed effects, and university-specific linear trends. The sample contains 141 universities and spans the biennial intervals from 1994 to 2016. All regressors are lagged 2 years. To reflect the biennial structure of our sample, the counts of patents and R&D spending are derived by combining values from odd years with those from even years, from 1993 to 2016. In column 3, I scale the main regressor, international graduate STEM student share by multiplying it by 1000. In column 5 where the dependent variable is binary, I scale the regressors by dividing them by 1000.

Chapter 6:
CONCLUSION

In this thesis, I have investigated the effects of international graduate students in STEM fields on the quantity and quality of innovative outputs of US universities in biennial periods from 1994 to 2016. The results show that impact of international graduate students have a positive and significant on the patent counts, but not on the patent citations. There are at least two reasons that can explain the positive impact of international graduate students on patent production of US universities. First, U.S. universities' STEM departments typically admit high-quality international graduate students who often rank among the top students in their home countries. This admission process ensures a pool of talented individuals capable of generating innovative ideas, thus positively impacting patent production. Second, visa restrictions limit international students' off-campus employment opportunities compared to the domestic students, potentially encouraging them to focus on their academic career and increase their research productivity.

As Chellaraj et al. (2008) demonstrate that the international graduate students also positively impact patents awarded to non-university institutions, indicating their contribution extends beyond academia. Therefore, international graduate students enhance the overall innovative capacity of the US. As the collaboration between universities and firms in research intensifies, international graduate students in STEM fields are likely to have a greater effect on the US innovation. Consequently, their impact on the US economy is likely to grow. All in all, policymakers should carefully consider the potential consequences of regulations affecting international graduate student immigration. Given their contributions to research and patent production, restrictive policies may harm the innovation production capability of the US. Therefore, it is crucial to create an attractive environment for top global talent in graduate education.

BIBLIOGRAPHY

[Autor et al., 2013] Autor, D.H., Dorn D., & Hanson, G.H. (2013). The China syndrome: Local labor market impacts of import competition in the United States. *American Economic Review 103*, 2121-2168.

[Babina et al., 2004] Babina, T., Alex X. H., Howell, S. T., Perlman E. R., & Staudt J. (2023). Cutting the innovation engine: How federal funding shocks affect university patenting, entrepreneurship, and publications. *The Quarterly Journal of Economics*, 895-954, https://doi.org/10.1093/qje/qjac046

[Barnow et al., 1981] Barnow, B., Glen C., and Goldberger A. (1981). Selection on observables. *Evaluation Studies Review Annual 5*(1), 43-59.

[Bartik, 1991] Bartik, T.J. (1991). Who benefits from state and local economic development policies? *W.E. Upjohn Institute for Employment Research.*

[Borjas, 1994] Hall, B. (2004). The economics of immigration, *Journal of Economic Literature*, vol. 32, pp. 1667–717.

[Borjas et al., 1997] Borjas, G., Freeman, R., & Katz, L. (1997). How much do immigration and trade affect labor market outcomes. Brookings Papers on Economic Activity, 1, 1-85.

[Cameron and Trivedi, 2009] Cameron, A.C. and Trivedi, P.K. (2009) Microeconometrics Using Stata. Stata Press, Texas.

[Card, 1990] Card, D. (1990). The impact of the Mariel boatlift on the Miami labor market, *Industrial and Labor Relation Review 43, 245-57.*

[Card, 2009] Card, D. (2009). Immigration and inequality. *American Economic Review: Papers & Proceedings 99*, 1-21.

[Chellaraj et al., 2008] Chellaraj, G., Maskus, K. E., & Mattoo, A. (2008). The contribution of international graduate students to US innovation. *Review of International Economics*, 16, 444-462. https://doi.org/10.1111/j.1467-9396.2007.00714.x

[Goldsmith-Pinkham et al., 2020] Goldsmith-Pinkham, P., Sorkin I., and Swift, H. (2020). Bartik Instruments: What, When, Why, and How. *American Economic Review 110*(8), 2586-2624.

[Guimãres, 2008] Guimãres, P. (2008). The fixed effects negative binomial model revisited. *Economics Letters, Elsevier*, 99(1), 63-66.

[Gurmu et al., 2010] Gurmu, S., Black, G. C., & Stephan, P. E. (2010). The knowledge production function for university patenting. *Economic Inquiry, 48,* 192-213.

[Hall, 2004] Hall, B. (2004). Exploring the Patent Explosion. *Journal of Technology Transfer*, 30, 35-48.

[Hall et al., 2001] Hall, B., Jaffe, A., & Trajtenberg, M. (2001). The NBER patent citations data file: lessons, insight and methodological tools. NBER Working Paper.

[Hausman et al., 1984] Hausman, J., Bronwyn, H., and Griliches, Z. (1984). Econometric models for count data with an application to the patents-R&D relationship. *Econometrica 52*(4): 909-938.

[Hsu et al., 2021] Hsu, D. H., Hsu P., Zhou T., Ziedonis A.A. (2021). 'Benchmarking U.S. university patent value and commercialization efforts: a new approach', *Research Policy (50).*

[Hunt and Gauthier-Loiselle, 2010] Hunt, J., & Gauthier-Loiselle, M. (2010). How much does immigration boost innovation? *American Economic Journal: Macroeconomics 2(2), 31-56.*

[Jaffe and Trajtenberg, 2005] Jaffe, A. and Trajtenberg M. (2005). Patents, citations and innovations: a window on the knowledge economy, Cambridge, MA: MIT Press.

[Jaravel, 2019] Jaravel, X. (2019). The unequal gains from product innovation: Evidence from the US retail sector. *The Quarterly Journal of Economics 134*, 715-783.

[Kahn, 2010] Kahn, L. (2010). The long-term labor market consequences of graduating from college in a bad economy. *Labour Economics 17*(2), 303-316.

[Kerr and Lincoln, 2010] Kerr, W. R., & Lincoln, W. F. (2010). The Supply Side of Innovation: H-1B Visa Reforms and U.S. Ethnic Invention. *Journal of Labor Economics*, *28*(3), 473–508. https://doi.org/10.1086/651934

[Kim et al., 2005] Kim, J., Lee S.J., and Marschke G. (2005). The influence of university research on industrial innovation (NBER Working Paper No.11447). National Bureau of Economic Research.

[Lerner and Seru, 2021] Lerner, J., & Seru, A. (2021). The use and misuse of patent data: issues for finance and beyond. T*he Review of Financial Studies*.

[Lin and Wooldridge, 2019] Lin, W., & Wooldridge, J.M. (2019). Testing and Correcting for Endogeneity in Nonlinear Unobserved Effects Models. In M. Tsionas (Ed.), *Panel Data Econometrics: Theory* (pp. 21-43).

[Monath et al., 2021] Monath, N. J., Jones C., Madhavan, S. (2021). 'PatentsView disambiguating inventors, assignees, and locations', *American Instutes for Research*.

[Mowery et al., 2001] Mowery, D., Nelson R., Sampat B., and Ziedonis A. (2001). The growth of patenting and licensing by US universities: an assessment of the effects of the Bayh–Dole Act of 1980. *Research Policy 30*, 99–119.

[Oberfield and Raval, 2021] Oberfield, E. & Raval D. (2021). Micro data and macro technology. *Econometrica* 89, 703-732.

[Pakes and Griliches, 1980] Pakes A. & Griliches Z. (1980). Patents and R&D at the firm level: A first report. *Economics Letters, Elsevier, vol. 5(4),* 377-381.

[Popp et al., 2003] Popp, D., Juhl T., Johnson D. (2003). 'Time in purgatory: determinants of the grant lag for U.S. patent applications', NBER Working Paper.

[Stephan and Levin, 1992] Stephan, P.E., and S.G. Levin (1992). Striking the mother lode in science: The importance of age, place, and time. *New York: Oxford University Press*.

[Stuen et al., 2012] Stuen, E. T., Mobarak, A. M., & Maskus, K. E. (2012). Skilled immigration and innovation: Evidence from enrolment fluctuations in US doctoral programmes. *The Economic Journal*, *122*(565), 1143-1176. https://doi.org/10.1111/j.1468-0297.2012.02543.x

[Thursby and Kemp, 2002] Thursby, J. and Kemp S. (2002). Growth and productive efficiency of university intellectual property licensing. *Research Policy* 31, 109–24.

[Wooldridge, 1999] Wooldridge, J.M., (1999). Distribution-Free Estimation of Some Nonlinear Panel Models. *Journal of Econometrics 90*, 77-97.

[Wooldridge, 2015] Wooldridge, J.M., (2015). Control function methods in applied econometrics. *Journal of Human Resources 50*(2), 420-445.