The Impact of Immigration on Green Technology Innovation in U.S. MSAs

Salvatore Viola[*]

Advisors: Rosina Moreno[†] & Ernest Miguelez[‡]

Universitat de Barcelona – UB School of Economics

Master's Thesis

June 15, 2023

Abstract

Climate change poses a persistent and worsening threat to humanity. A key component which will contribute to the transition to sustainable practices is the development and implementation of green technology (GT). Significant research has been conducted investigating the determinants of innovation. One central determinant is immigration and location. While the literature investigating immigration, innovation and location is time-tested and there exists a growing body of studies relevant to the determinants of GT, little has been done to understand the impact of immigration on GT in particular. This paper motivates the construction of a unique cross-sectional dataset using shares of foreign workers and inventors, and patent counts. Negative binomial estimations help to investigate shares of foreign workers as a potential determinant of GT at the U.S. MSA level.

Keywords: Immigration, Green Technology, Innovation, Patents, Negative Binomial Regression, U.S., Metropolitan Statistical Areas

JEL: Q55, R11, F22

[*] salvatorefviola@gmail.com, Universitat de Barcelona
[†] rmoreno@ub.edu, Universitat de Barcelona
[‡] ernest.miguelez@u-bordeaux.fr, Bordeaux Sciences Economiques

1 Introduction

Climate change is a persistent and worsening challenge which our society faces, one that threatens global welfare. The imperative to combat it grows greater with each passing year. Experts believe that the world now possesses sufficient capital, policy instruments, and scientific knowledge to halve carbon emission by 2030 (IPCC, 2022). However, the consequences of inaction loom and may be calamitous (Haines and Patz, 2004; McMichael et al., 2006). Solutions must therefore involve a dynamic shift to a low-carbon economy through innovation and the implementation of low-carbon technologies and other sustainable practices. One contribution to this transition will be made by the development and improvement of green technologies (GT), those technologies intended to mitigate or reverse the effects of human activity on the environment. GT is also known by several other names such as environmental innovation (Durán-Romero and Urraca-Ruiz, 2015). Studies by Favot et al. (2023), Desheng et al. (2021), and Chen et al. (2021) explain how green innovation can be assessed by a variety of different methods including research and development expenditures, green total factor productivity and, most importantly for present study, green patents.

Patents have long been the foci of economic studies including those related to topics of immigration, and to what extent international migration in particular affects innovation. The mechanisms suggest that influxes in high skilled labor have the potential to spur innovation in the locations which receive it. Additionally, it is supposed that skilled migration may also import new knowledge onto host societies, which also has implications for patenting activity. GT patents are a particular case which is beginning to become the focus of a number of other studies. Recently, investigative efforts have been made to better classify GT and study its determinants. Papers such as Favot et al. (2023) have succeeded in selecting classification measures for GT patents which identify the greatest number of patents closest to GT in sample datasets using methodology developed by the Organization for Economic Co-operation and Development (OECD). The relationship between immigration and GT, however, is still somewhat understudied. Some evidence suggests that the development of environmental technologies requires a more complex and diversified knowledge set due to their novelty. Considering the potential that migrants have to introduce new knowledge to their destination, there is reason to believe that there may be a connection between the two.

The USA is one of the global leaders in innovation and has a rich history of migration, one which has long attracted skilled workers. Studies of this kind may be undertaken at the country level. However, previous research indicates that certain geographical areas like cities or metropolitan statistical areas (MSAs) are not only hubs for immigration, but also for innovation and patenting activity. Therefore, the present study examines the case of migration and innovation in U.S. MSAs. In addition to the country's characteristics which make it an appropriate location for the purpose of this study, government databases like the U.S. Census Bureau, the Integrated Public Use Microdata Series (IPUMS), and the U.S. Patent and Trademark Office provide ample data for these purposes. This study exploits the IPUMS USA dataset for the year 2006 and U.S. patent data from the USPTO's PatentsView database in order to investigate the relationship between shares of international immigrants and GT patents.

This thesis contributes to the literature in two ways: first, it considers an understudied relationship between shares of skilled immigrants and inventors and GT at a relevant geographical scale in the U.S. Second, it employs a unique dataset which relies on an uncommon aspect of PCT data which identifies inventor nationality and exploits the most recent and appropriate methods of GT identification in patent data. The results of the present study indicate that shares of highly skilled migrants are positively correlated with overall counts of GT patents in U.S. MSAs in 2006. This is also true in the cases of certain subsets of this classification such as Y02 patents, "which […] can be considered as countering the effects of climate change, namely technologies or applications which can decrease greenhouse gases (GHG) emission or remove (and store) GHG from the atmosphere" (Angelucci, et al., 2018, p. S86). While the shares of skilled migrants retain some explanatory power in this analysis, the inclusion of variables which capture the shares of foreign inventors appear to be more related to GT patents. Not only are the share of skilled migrants and foreign inventors able to explain changes in counts of patents, but they also have implications for the technological specialization of MSA as higher shares of foreign inventors were positively correlated with indicators of revealed technological advantage (RTA) in Y02 technologies. The results have important implications for the relationship between shares of skilled foreigners and inventors and the generation of GT.

The rest of this thesis is organized as follows: Section II summarizes relevant findings of related papers, Section III provides an in-depth description of the data sources used and variables created for the purpose of this study, Section IV describes the methodology used in this analysis, Section V presents and discusses the results of this analysis, and Section VI concludes this thesis.

2 Literature Review

Remedying climate change will depend on the engagement and efforts from a variety of industries and fields which may be more directly implicated. That being said, economics will play a crucial role in this transition. In particular, the study of innovation economics will be at the forefront. While their paper principally addresses the effect of intellectual property protection in technological implementation and adoption, Hall and Helmers (2010) claim that "global climate change mitigation will require the development and diffusion of a large number and variety of new technologies" (p. 2), including green technologies (GT). Aghion et al (2009) argue that the existence of technologies to combat climate change are being treated as "given" and that targeted innovation in green technology is crucial to global movements towards sustainability.

Despite the displayed relevance of GT in combating climate change, why is it important that its determinants be studied? An important characteristic of GT, as previously mentioned, is that its development requires a specific set of capabilities, which appear to be far different from traditional knowledge bases of industries (Orsatti et al., 2020; Perruchas et al., 2020). Furthermore, GT are more complex, novel, pervasive and impactful than the non-green ones (Barbieri et al., 2020), and require high diversity of competences (Zeppini and van der Bergh, 2011). While it has been shown that individual teams' creative recombination of capabilities boosts GT (Orsatti et al., 2020), it has also been found that high density of related regional technologies boost GT as well. (Montresor & Quatraro, 2020). Considering the complicated nature of these technologies and their apparent relevance, it is crucial that the processes are understood which contribute to their production, specifically in areas which might supply this high density of related regional technologies such as MSAs.

Immigration has long been a central topic in the study of certain economic outcomes and the study of innovation is no exception. There exists significant evidence that immigration is strongly connected with innovation and even provokes it since foreigners may be attracted to universities and may increase collaboration between academic and non-academic groups (Hunt and Gauthier-Loiselle, 2010; Chellaraj et al., 2008; Stephan and Levin, 2001). Such studies have also motivated the investigation of the mechanisms through which this interaction occurs. Lissoni (2018), for example, finds that immigrants boost local innovation through the dissemination of immigrants' technical and scientific knowledge into their host societies. Additionally, many researchers have studied this effect using historical examples such as the migration of Soviet scientists in Ganguli (2015), the Huguenots in Hornung (2014), and the migration of German Jews in Moser, et al. (2014). Recalling that GT requires a specific and diverse set of skills, does immigration affect this as well? A number of authors have investigated the effect of immigration on local skill composition and diversity (Alesina et al., 2016; Kemeny, 2017; Ozgen, et al., 2012). The general finding of this literature is that migration increases the diversity of both culture and knowledge of the cities in which they accept employment.

Another body of literature then takes the initiative to explore the effects of this diversification. Authors such as Bathelt et al. (2004) and Morrison et al. (2013) have shown that skilled immigration can create channels to access new and nonredundant knowledge (which could manifest as GT). Similarly, other papers have shown that skilled immigration can cause technological diversification, specifically relating to those technologies of the home countries of the immigrants (Miguelez & Morrison, 2022). Bahar, Rapoport, Turati (2019), for example, find that birthplace diversity is positively associated with economic complexity because immigrants expand the set of skills to which a country has access. The review of this literature thus reveals an interesting link: if immigrants in general may be a driver of innovation and environmental technologies may require a more diverse set of competencies which may be supplied by non-native workers, does immigration drive innovation in GT in particular?

3 Data and Descriptives

3.1 Data and Variables

The data for this paper originate from a number of different sources including the USPTO, PCT, IPUMS USA, the United States Census Bureau's Metropolitan and Micropolitan Statistical Areas Population Totals and Cartographic Boundary Files databases, and the National Center for Education Statistics university location files. Via the PatentsView website, the USPTO offers several disaggregated patent level datasets. These include datasets containing inventor characteristics, firm characteristics, locations of the patent and corresponding patent classifications. The IPUMS USA is a database providing microdata on residents within the U.S. gathered from previous U.S. censuses. Relevant variables include individual levels of education, age, gender, labor force, nativity and geographic indicators which can be used to aggregate the data at the MSA level in the U.S. Since the IPUMS USA draws samples from larger census data, the number of available years is restricted to five-year intervals. For that reason, the present study will exploit data from IPUMS USA from the year 2006. Previous analysis using other U.S. microdata bases such as the IPUMS CPS which contains observations at the individual level on a year (and more recently, monthly basis) yielded variable counts which were relatively volatile and were not considered to be statistically adequate for this type of analysis.

The two principal variables of my study are GT innovation and skilled immigration. While all of the necessary information is contained in the various USPTO PatentsView datasets, a methodology must be adopted in order to merge them. A common key between these datasets is the patent ID numbers. These identification numbers can be used to combine a number of the datasets; however, further work is required to match the patent data based on the ambiguated location identifier. Therefore, a combination of both of the patent ID numbers and the ambiguous location data can be used to combine the numerous datasets into one master patent data set which contains variables for the patent ID, the inventor names, numerous variables relating to the patent location, including the latitude and longitude of the assignee, the assignee name and type, and both the IPC and CPC[1] classification codes.

The next challenge is to identify the patents within the datasets which are considered GT. According to Favot et al. (2023) referencing Kraus et al. (2020) "Green innovation (GI) refers to the innovation in technology applied to minimize wastage, global warming, use of water, air pollution, use of coal, oil, electricity, and conserving energy" (p. 1). While recent literature relating to GT innovation has developed and applied a number of different methodologies in order to identify GT patents, Favot et al. (2023) recommend using a combination of CPC and IPC classification codes which they claim identifies the largest number of green technology patents in their samples. This combination of CPC and IPC codes has been developed by the OECD and named the ENV-TECH classification method which is regularly updated to reflect current development in the identification methodologies (Haščič & Migotto, 2015). Through the ENV-TECH methodology, a list of corresponding codes can be obtained and matched with the corresponding CPC and IPC codes in my master patent dataset. Accordingly, any patent which has been granted and lists a code which is also contained in the ENV-TECH identification methodology as developed by the OECD and described in Favot, et al. (2023) is considered GT. This subsequently allows for the

[1] According to a summary provided by the Queen's University Library (Queen's), the Cooperative Patent Classification (CPC) is a patent classification system developed by the EPO and USPTO that contains approximately 200,000 subgroups and the International Patent Classification (IPC) is a hierarchical classification system consisting of about 70,000 subgroups. Generally, they are used to identify which type of specific technological category a patent belongs to. These classifications in turn have been used to identify which technologies may be considered environmentally related.

creation of a dichotomous variable, GT, which takes the value of one for those patents that have a matching CPC or IPC code. Hereafter, the dataset can be reduced to those patents which are located within the U.S. and were granted between, during or after 2006.

A concern regarding this newly created master patent dataset is that it does not contain the MSA location of each of the patents. Recall that the patents are, however, geo-localized. Therefore, using MSA shapefiles for the U.S. patent locations can be identified and assigned. Once the patents have been successfully assigned, total patent counts can be created which reflect the total count of patents in a given MSA in a given year in addition to the counts of GT patents. While these variables alone may capture some of the concentration of non-GT and GT patents, there exists great diversity within the GT innovation itself. Therefore, I also create variables which correspond to the counts of those patents identified by the Y02 tagging scheme in addition to each of the 8 subcategories of this identification method.

Y02 Technology Categories (OECD ENV-TECH)[2]	
Y02A	Climate change adaptation technologies
Y02B	Climate change mitigation technologies related to buildings
Y02C	Capture, storage, sequestration or disposal of greenhouse gases
Y02D	Climate change mitigation in information and communication technologies
Y02E	Climate change mitigation technologies related to energy generation, transmission or distribution
Y02P	Climate change mitigation technologies in the production or processing of goods
Y02T	Climate change mitigation technologies related to transportation
Y02W	Climate change mitigation technologies related to wastewater treatment or waste management

[2] Full list available at https://stats.oecd.org/OECDStat_Metadata/ShowMetadata.ashx?DataSet=PAT_IND

The next principal dataset for this research is the IPUMS USA microdata. As previously noted, the census data is only available for a number of years in the U.S. Therefore, the year 2006 will be included in the present study. The IPUMS USA contains variables which locate its respondents, assign them workforce status, and identify their level of education and nativity. From these variables, individuals can be located and coded as workers possessing various levels of education and nativity. Workers are identified as those individuals who have a positive workforce participation status. For the purpose of this study, two levels of education are defined: skilled workers, or those individuals who possess a four-year college degree or higher, and high skilled workers, or those individuals who possess at least a doctorate degree. As will become apparent in the subsequent analysis, only highly skilled workers will be considered. Next, workers can be identified as foreign or native based on their nativity. Foreign workers are identified as those individuals who reported birth places outside of the U.S. Conversely, native workers are those individuals who reported a birthplace in the U.S. The identification of these individual characteristics allows for the calculation of the share of the ranging levels of education that are foreign or native born by the U.S. MSA. For example, the variable of the share of skilled workers that are foreign is calculated as the number of foreigners with at least a four-year college degree in a given MSA divided by the total number of workers in the same MSA that have at least a four-year college degree. The same methodology is also applied to the other skill levels.

While data alone may yield interesting results to identify certain clusters of innovation or GT innovation, some controls must be included in the data. The primary control for the present study is the shares of highly skilled human capital by MSA. Data from IPUMS USA also allows for all workers with a doctorate degree or above to be identified and used to calculate these shares. Another control which I propose is population. The U.S. Census Bureau publishes yearly datasets which contain population estimates by varying geographical denominations. For the purpose of this study, population estimates for the year 2006 at the MSA level are extracted and applied to the data. In addition to population, I also propose controlling for the presence of universities in MSAs. The Institute of Education Sciences National Center for Education

Statistics has been publishing yearly since 2015 the location of all postsecondary educational institutions in the U.S. From this data, the total count of universities can be calculated at the MSA level. Similar to IPUMS USA, the postsecondary institution data contains CBSA codes which can be used to match the counts of universities with the master data. Some research has suggested that universities are responsible for a portion of overall innovative activity (National Research Council, 2012). Therefore, I considered it relevant to control for the potential impact the universities may be having on local innovative activity. An important consideration is that this variable is from 2015 and not the year of study 2006. Since 2006, however, only a handful of new universities have been established in the U.S. and these observations can be removed from the data to accurately reflect the counts in 2006.

The aim of the introduction of the share of foreigners in the worker force is intended to capture the impact that such individuals may have on local innovative activity. The main limitation with the aforementioned variable is that it captures the proportion of all skilled foreign individuals but does not necessarily accurately reflect the proportion of skilled workers implicated in the innovation process, that is to say, inventors themselves. Therefore, this thesis also exploits data from the World Intellectual Property Organization (WIPO) IPSTATS database and looks at inventors residing in one U.S. MSA with patents applied through the Patent Cooperation Treaty (PCT), administered by WIPO. Under the PCT, data for inventor residence and nationality[3] can be obtained, which allows for the creation of another variable which captures the share of foreign inventors at the MSA level. This data spans the years of 2004 to 2006 in order to capture the lagged effects that the presence of certain types of inventors may have on innovation. These values can be added to the master dataset, merging them with the corresponding MSA codes.

While patent counts alone, when controlling for population, can be indicative of higher levels of innovation in given MSAs, we also want to analyze to what extent the presence of skilled foreigners and foreign

[3] The availability of patent data which contains corresponding inventor nationality is quite unique. It is only available through the PCT for those patents which filed for protection in the U.S. between the years 1990 and 2010. Certain patent application procedures in the U.S. during this period apparently required that inventors be listed as applicants as well, and that their nationality and residence be recorded (Miguelez & Fink, 2013).

workers may be related to the specialization pattern of MSAs. To this end, I propose the use of a measure of revealed technological advantage (RTA). In the case of GT patents, I create a variable which captures if a given MSA has a revealed technological advantage in such a group of technologies. The same for all Y02 tagged technologies and the same for each of the 8 individual classifications of Y02 technologies. The purpose of this variable is to measure to what extent the shares of skilled workers and inventors of different origins may be related to technological differentiation and diversity in a given MSA. As mentioned, some of the previous literature (Bathelt et al., 2004; Morrison et al., 2013; Miguelez & Morrison, 2022) has found that the presence of foreign inventors may promote diversification into new areas of technology and the creation of new technology overall.

In addition to the individual measures of RTA for separate types of GT, I propose another variable which measures the extent to which this effect takes place. Therefore, I create a variable which is a count of the sum of the previously described RTA indicators. That is the variable could take a value between 0 and 8 depending on the number of different RTAs that a given MSA has in 2007. While certain MSA characteristics may explain whether they have an RTA in a specific type of technologies, this variable aims to capture the extent to which technological diversification takes place based on Y02 patents.

3.2 Descriptive Statistics

Table 1 displays the descriptive statistics for the main variables of interest including patent counts and the control variables. As the table indicates, patent data is used for a total of 373 U.S. MSAs in the year of 2007. The average number of patents per year per MSA is around 255 whereas the average count of patents classified as GT is about 29. Surprisingly, the number of those GT patents which are also classified as Y02 technologies is around 28, which is quite similar to the average of all GT patents. This would indicate that the vast majority of all patents classified as GT patents are identified by the Y02 tagging scheme, which is much different from the findings of Favot et al. (2023) who found that this percentage was much lower. Accordingly, the counts of patents belonging to the various subcategories of Y02 patents are much lower, averaging around 8 at the most and less than 1 at the lowest. All of the patent categories show a large amount of variation across MSAs. This is shown in part by the standard deviations, but also by the maximum and minimum values. While in 2007 some MSAs produce only a single patent, others have nearly 10,000. The same variation is true for GT patents which are not even present in certain MSAs and account for thousands of the patents in others. As will become clear, some of this variation is likely caused by scale effects and will likely have to be accounted for in the regressions by controlling for population. Another method to address the issue of high variation would be to drop observations identified as outliers. However, count data is somewhat unique as it may present a non-normal distribution and the task of dropping outliers may not be appropriate due to the large number of true zeros in the dataset.

Table 1: Patent Counts and Controls

Variable	Obs	Mean	Std. Dev.	Min	Max
Total Patents	373	254.98	819.78	1	9787
GT	373	29.48	84.97	0	1000
Y02	373	28.09	80.82	0	829
Y02A	373	4.11	14.31	0	119
Y02B	373	2.6	7.95	0	65
Y02C	373	.36	1.64	0	18
Y02D	373	3.49	15.53	0	213
Y02E	373	6.94	19.59	0	144
Y02P	373	7.05	22.79	0	211
Y02T	373	5.92	36.18	0	642
Y02W	373	.97	3.13	0	36
Universities	373	17.18	36.32	1	418
Population	373	260459.18	661218.35	105	9700359

Table 2 shows the descriptive statistics for the main explanatory variables of interest. An important figure from the table is the number of observations. In contrast to the patent data, shares of foreign workers and inventors are only available for the subset of the MSAs for which patent data is available. Nevertheless, the entire sample will still examine a total of 239 MSAs. On average, according to the IPUMS USA data, in U.S. MSAs skilled foreigners represent around 10% of all skilled workers. The ratio for highly skilled foreign workers of all highly skilled workers in a given MSA is 12%. This entails that highly skilled foreigners make up a slightly large proportion of all highly skilled workers in the same MSA than for just skilled workers. The variables for the shares of foreign inventors displays a similar ratio, where foreigners account for around 11% of all inventors.

Table 2: Shares of Skilled Foreigners and Human Capital

Variable	Obs	Mean	Std. Dev.	Min	Max
Share of Skilled Foreigners	239	.1	.07	0	.46
Share of High Skilled Foreigners	239	.12	.08	0	.52
Share of Foreign Inventors	239	.11	.08	0	.58
Share of High Skilled Human Capital	239	.1	.04	.03	.26

Table 3: Revealed Technological Advantage (RTA) of Patent Types

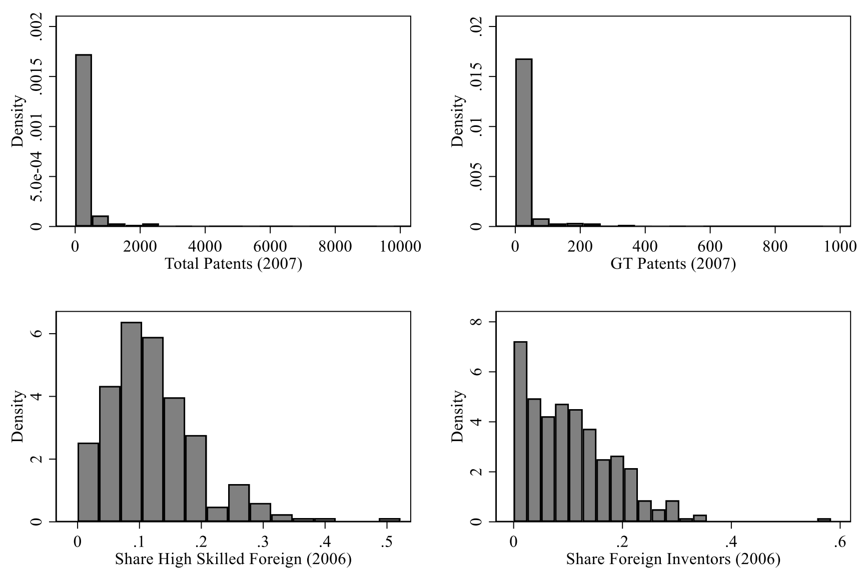
Variable	Obs	Mean	Std. Dev.	Min	Max
RTA Y02	373	.24	.43	0	1
RTA GT	373	.25	.43	0	1
RTA Y02A	373	.2	.4	0	1
RTA Y02B	373	.18	.38	0	1
RTA Y02C	373	.06	.24	0	1
RTA Y02D	373	.09	.28	0	1
RTA Y02E	373	.21	.41	0	1
RTA Y02P	373	.2	.4	0	1
RTA Y02T	373	.13	.34	0	1
RTA Y02W	373	.17	.38	0	1
Cum. RTA	373	1.24	1.26	0	6

Table 3 provides descriptive statistics on the set of previously described RTA variables. They are present for the entire sample of patents and take values according to their description. A mean value of 0.24 would indicate that 24% of the MSAs have a RTA in this type of technology. We can therefore discern that a greater number of MSAs have an RTA in technologies like GT overall, Y02 patents, and Y02P (climate change mitigation technologies in the production or processing of goods). The final variable in the table, Cumulative RTA, is the variable which captures the sum of different RTAs. The mean of this variable is just above 1, suggesting that, on average, MSAs have an RTA in at least one of the 8 categories of Y02 technology, but none displays patents in all 8 in 2007.

The raw correlations of all of the variables are presented in Table A1 in the appendix for the year 2007 using the lags of the shares of foreign workers and inventors. Of primary interest are the correlations between the share of skilled and highly skilled foreign workers and inventors with the set of patent counts. In all cases relating to the shares of foreign workers of all skill levels and foreign inventors, the correlation coefficient is positive and of a reasonable magnitude. In particular, the correlation coefficient on the share of skilled foreigners with total patents is 0.55 and that of highly skilled foreigners is 0.53. Alternatively, the coefficient on the share of highly skilled human capital of the entire skilled workforce is 0.46. As previously mentioned, while the share of skilled workers in the labor force may be correlated with levels of innovation, we should also consider the particular impact that inventors in particular have on this. Looking at the correlation coefficient on the share of foreign inventors we observe a similar trend. That is, the correlation on the share of foreign inventors is 0.38. Lastly, the correlation coefficients on the university counts and population are 0.63 and 0.38, respectively.

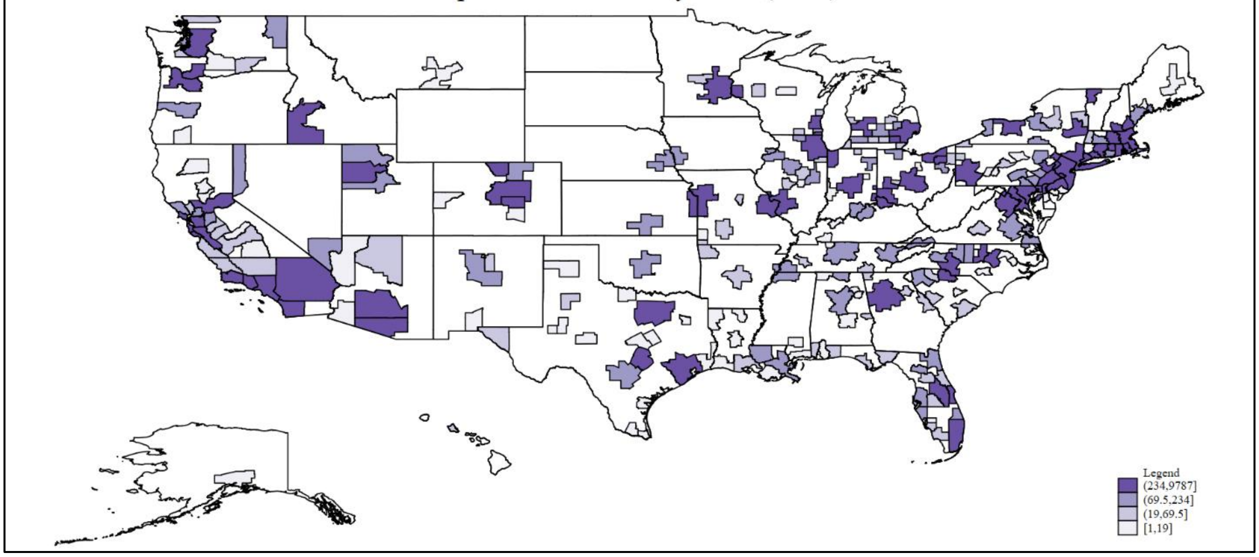
Below, Figure 1 contains graphs of the distributions of total and GT patent counts for the year 2007 and shares of skilled foreigners and foreign inventors for the year 2006. While the variables for the shares of skilled foreigners and foreign inventors follow slightly right-skewed distributions, the counts of Total and GT patents are extremely right-skewed, likely as a result of a number of low-count MSAs in the case of total patents and a number of MSAs with zero GT patents. As expected with count data, they

Figure 1: Distribution of Main Variables

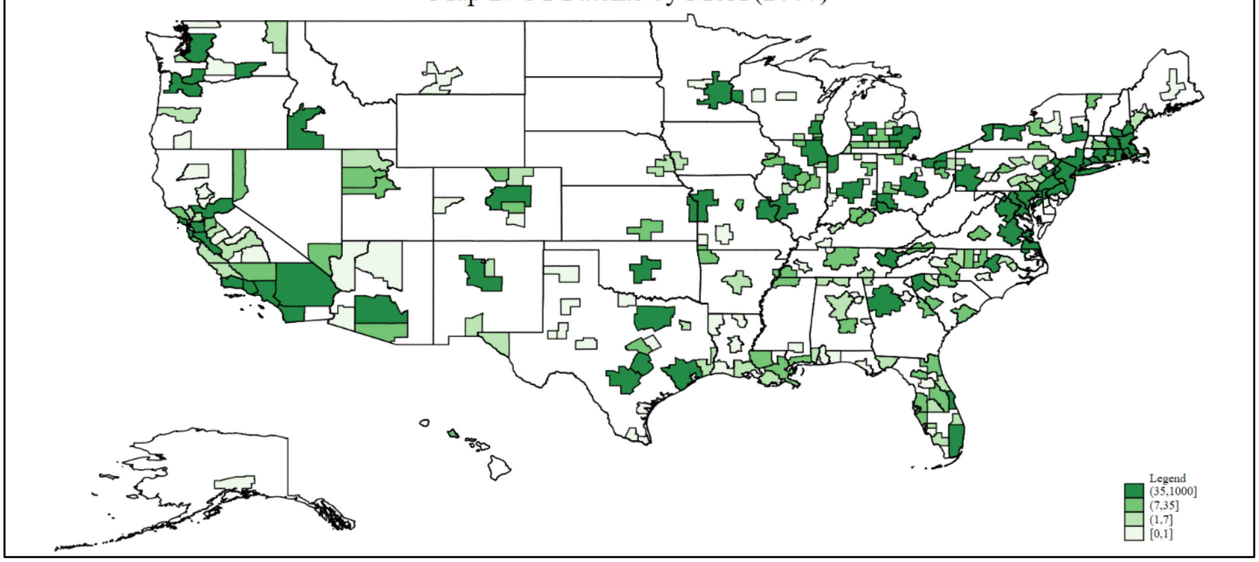


Maps 1-4 present choropleth maps of the counts of patents and share of foreign workers and inventors by MSA. Clearly, there is much spatial heterogeneity in the case of the central U.S. On the other hand, the coastal U.S. shows some strong heterogeneity, especially in the case of patent counts. These trends appear to coincide with locations in which there are a large number of universities such as eastern Massachusetts, southern Maine, and Connecticut, or larger tech industries such as in southern California. While a spatial visualization of the data is useful to understand its dispersion, the study does not employ spatial econometric methods and requires other methodologies to uncover the determinants of GT at the U.S. MSA level.

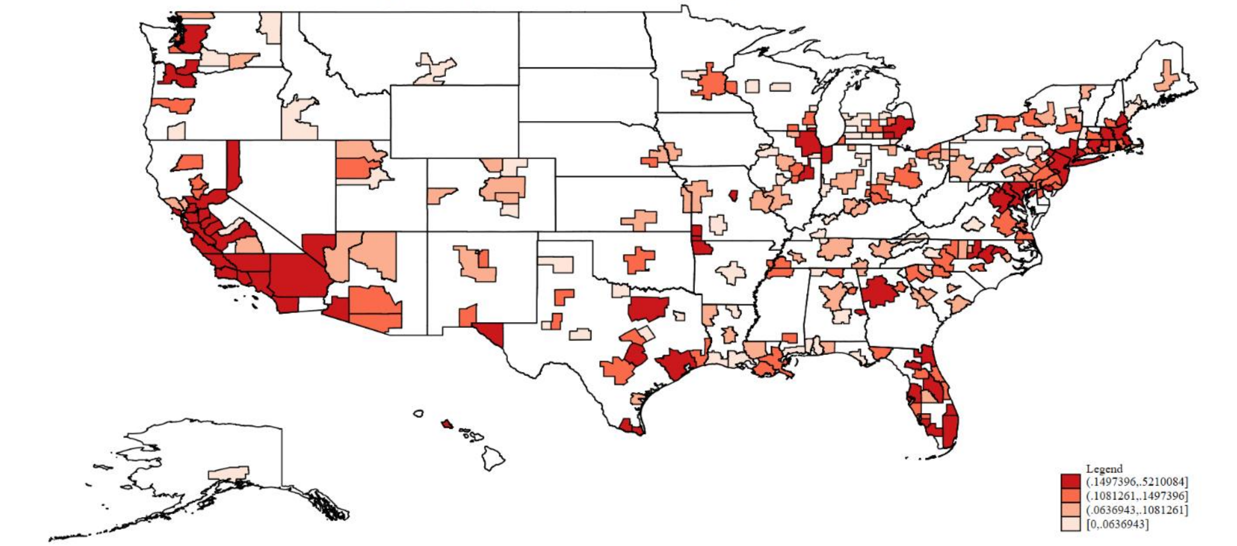
Map 1: Total Patents by MSA (2007)



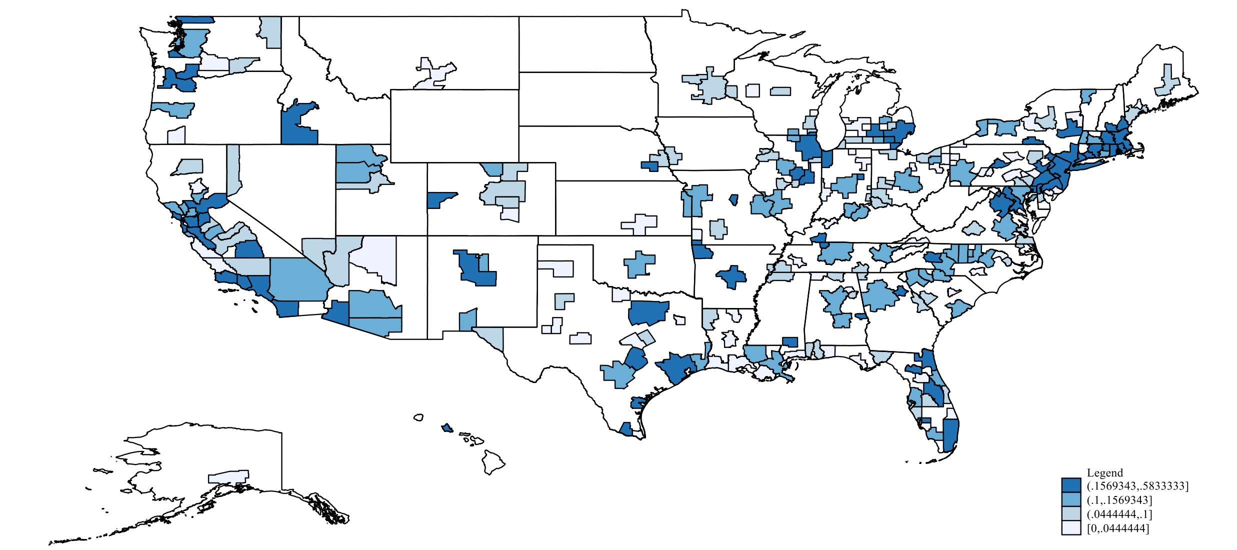
Map 2: GT Patents by MSA (2007)



Map 3: High Skilled Foreigners by MSA (2006)



Map 4: Foreign Inventors by MSA (2006)



4 Methodology

The primary empirical approaches of this paper include negative binomial and logistic regressions. Count data in general do not follow a normal distribution and are not appropriately estimated by ordinary least squares (OLS). Typically, Poisson estimations are performed in this case, however, they may become inappropriate as dispersion increases in the dependent variable. Therefore, in order to justify the use of the negative binomial estimation method in place of OLS, this thesis performs series of regressions and tests on adaptations of the following set of regressions:

(1) $$Pat_{i,t} = \beta_0 + \beta_1\, ShareSkill_{i,t-1} + \beta_2\, ShareInv_{i,t-1} + \beta_3\, X_{i,t} + \epsilon_i$$

(2) $$GT_{i,t} = \beta_0 + \beta_1\, ShareSkill_{i,t-1} + \beta_2\, ShareInv_{i,t-1} + \beta_3\, NonGT_{i,t} + \beta_4\, X_{i,t} + \epsilon_i$$

(3) $$Y02_{i,t} = \beta_0 + \beta_1\, ShareSkill_{i,t-1} + \beta_2\, ShareInv_{i,t-1} + \beta_3\, GT_{i,t} + \beta_4\, NonGT_{i,t} + \beta_5\, X_{i,t} + \epsilon_i$$

where $Pat_{i,t}$ is the total count of patents in MSA i in year t, $ShareSkill_{i,t}$ is the proportion of the skilled or highly skilled workforce that is foreign, $ShareInv_{i,t}$ is the proportion of inventors in a given MSA in a given year that are foreign, X is a vector of control variables including population, the count of universities and the stock of highly skilled human capital, and Ɛ is the error term. The base year, t, of the regression is 2007. Therefore, as previously indicated, the lagged shares of foreign workers and inventors are used.

According to variable inflation factors (VIF) in Table A4 in the appendix, we can conclude that multicollinearity is not of concern in our case. Additionally, the residuals displayed heteroskedasticity as displayed in Figures A1 and A2 in the appendix. Even though constant variance of the residuals is not an assumption of the models used in this paper, the following estimations are conducted correcting for heteroskedasticity using clustered standard errors at the MSA level.

Due to the high dispersion of the total and GT patent counts, there is some concern that a Poisson estimation may not be entirely appropriate. One measure of overdispersion in the data can be constructed using the dispersion index which is calculated as the variance over the mean. Any value of this index which is above one indicates that the data are overdispersed and consequently violates the assumption of the Poisson model of a constant mean-variance relationship. In the case of GT patents, this index is well over 200, indicating that the data are overdispersed and would be better modeled by a negative binomial regression. Additionally, the performance of a chi-squared goodness of fit tests on all of the negative binomial estimations from Tables A5 and A6 in the appendix result in a highly statistically significant test statistic which confirms that the count data is not well fit by the Poisson model and that we should prefer a negative binomial estimation.

5 Results

Tables 4 and 5 present the estimations of equations (1) and (2) from the Methodology section which stagger the inclusion of the shares of highly skilled foreign workers and the share of foreign inventors, and the set of controls. In the case of total patent counts in Table 4, results in column (1) indicates that the share of highly skilled foreigners is responsible over the share of skilled foreigners for explaining changes in total patent counts, the inclusion of controls and the share of foreign inventors in columns (4) and (5) begin to explain this variance. Lastly, column (6) is estimated using robust standard errors clustered at the MSA level and shows that the significance of the regressors is preserved even with slightly larger standard errors. From now on, as commented before, we use robust standard errors.

Table 4: Negative Binomial Regressions - Total Patents

	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	Total	Total	Total	Total	Total	Total (Robust)
Skilled Foreign = L,	-3.344	0.659				
	(5.126)	(3.441)				
High Skilled Foreign = L,	16.85***	0.721	1.291	1.291	-0.0639	-0.0639
	(4.754)	(3.169)	(1.090)	(1.090)	(1.122)	(1.225)
High Skilled HC = L,		21.40***	21.38***	21.38***	18.28***	18.28***
		(2.548)	(2.548)	(2.548)	(2.599)	(2.984)
Universities		0.0245***	0.0244***	0.0244***	0.0239***	0.0239***
		(0.00339)	(0.00336)	(0.00336)	(0.00325)	(0.00526)
Population		1.34e-08	2.15e-08	2.15e-08	-2.81e-09	-2.81e-09
		(1.56e-07)	(1.50e-07)	(1.50e-07)	(1.48e-07)	(2.06e-07)
Foreign Inventors = L,					4.282***	4.282**
					(1.177)	(1.756)
Constant	3.671***	2.001***	2.000***	2.000***	1.999***	1.999***
	(0.194)	(0.225)	(0.225)	(0.225)	(0.217)	(0.240)
Observations	238	238	238	238	238	238

Standard errors in parentheses
*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table 5: Negative Binomial Regressions - GT Patents

	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	GT	GT	GT	GT	GT	GT (Robust)
Non-GT	0.000738***	0.000213**	0.000219*	0.000219*	0.000201*	0.000201*
	(0.000121)	(0.000106)	(0.000112)	(0.000112)	(0.000103)	(0.000114)
Skilled Foreign = L,	-11.12**	-8.956**				
	(4.649)	(4.126)				
High Skilled Foreign = L,	13.63***	7.748**	0.271	0.271	-2.050	-2.050
	(4.567)	(3.888)	(1.746)	(1.746)	(1.903)	(1.557)
High Skilled HC = L,		18.65***	19.16***	19.16***	15.04***	15.04***
		(3.464)	(3.456)	(3.456)	(3.659)	(3.489)
Universities		0.0227***	0.0234***	0.0234***	0.0243***	0.0243***
		(0.00515)	(0.00535)	(0.00535)	(0.00508)	(0.00691)
Population		-9.70e-08	-2.18e-07	-2.18e-07	-2.15e-07	-2.15e-07
		(1.95e-07)	(1.83e-07)	(1.83e-07)	(1.84e-07)	(2.07e-07)
Foreign Inventors = L,					4.673***	4.673**
					(1.767)	(1.856)
Constant	1.985***	0.240	0.228	0.228	0.377	0.377
	(0.231)	(0.363)	(0.366)	(0.366)	(0.360)	(0.344)
Observations	238	238	238	238	238	238

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 5 displays the same set of regressions on equation (2). While the shares of skilled and highly skilled foreign workers are significant in columns (1) and (2), the inclusion of the share of foreign inventors eliminates their relevance. Similarly, column (6) shows that the statistical significance of the regressors is preserved when estimating with corrected standard errors. One key observation across the regressions in Tables 4 and 5 is that shares of highly skilled human capital and the number of universities always seem to capture some explanatory power. Another is that the share of foreign inventors appears to be more relevant for explaining differences in both total and GT patents than the share of highly skilled foreigners alone.

Table 6 presents the estimation of a combination of equations (1), (2) and (3) both with and without the share of foreign inventors. In the case of total patents in equation (2), the share of foreign inventors is statistically significant and has a coefficient of 4.28. This indicates that for a one unit increase in the share of foreign inventors, the expected log count of the number of total patents increases by 4.28. Referring to

Table A7 in the appendix which lists the corresponding incident rate ratios shows that for every unit increase in the share of foreign inventors, the percent change of the incident rate of total patent count increases by 72.38%. For GT patents in column (4), a one unit increase in the share of foreign inventors, the expected log count of the number of GT patents increases by 4.67. Additionally, the number of universities are statistically significant and positive in all cases save the Y02 patents. While this makes intuitive sense, it is also an important result since the share of foreign inventors retains its statistical significance. This shows that, in the case of patents counts, the shares of foreign inventors are relevant even when controlling for overall highly skilled human capital.

Table A7 indicates that a one unit increase in the share of foreign inventors corresponds to a 107.1% increase in the percent change of the incident rate of GT patents. A relationship of similar magnitude is observed in the case of Y02 patent counts for which the magnitude of the coefficient is 4.08 and the incident rate ratio is about half that of GT patents. The predicted margins at the mean of the regressions of Table 6 are included in Table A9 in the appendix. Holding all other variables constant, the predicted number of GT patents by the share of foreign inventors is 183.6. In the case of high skilled capital, the predicted number of GT patents at its mean is 343.

Table 6: Negative Binomial Regressions by Patents Type with Inventors

	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	Total	Total	GT	GT	Y02	Y02
High Skilled Foreign = L,	1.291	-0.0639	0.271	-2.050	0.530	-1.208
	(1.196)	(1.225)	(1.461)	(1.557)	(1.265)	(1.275)
Foreign Inventors = L,		4.282**		4.673**		4.080***
		(1.756)		(1.856)		(1.410)
High Skilled HC = L,	21.38***	18.28***	19.16***	15.04***	15.41***	12.74***
	(2.729)	(2.984)	(3.171)	(3.489)	(2.615)	(2.620)
Universities	0.0244***	0.0239***	0.0234***	0.0243***	0.0101*	0.0101**
	(0.00525)	(0.00526)	(0.00739)	(0.00691)	(0.00517)	(0.00510)
Population	2.15e-08	-2.81e-09	-2.18e-07	-2.15e-07	1.03e-07	1.08e-07
	(2.00e-07)	(2.06e-07)	(2.01e-07)	(2.07e-07)	(2.37e-07)	(2.42e-07)
Non-GT			0.000219	0.000201*	-0.000260	-0.000268
			(0.000140)	(0.000114)	(0.000178)	(0.000164)
GT					0.0150***	0.0147***
					(0.00508)	(0.00509)
Constant	2.000***	1.999***	0.228	0.377	0.145	0.169
	(0.248)	(0.240)	(0.347)	(0.344)	(0.276)	(0.257)
Observations	238	238	238	238	238	238

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

While there appears to exist a clear relationship between the shares of foreign inventors and most of the control variables with total and GT patent counts, does this relationship exist for specific types of GT patents? Table 7 presents an adaptation of equation (3) in which the dependent variable is represented by a different Y02 technology patent count in each of the 8 regressions. Shares of highly skilled human capital are significant across 6 of 8 specifications and shares of foreign inventors are significant across 4. It would appear that the climate change mitigation technologies related to transportation (Y02T) and technologies related to wastewater treatment or waste management (Y02W) patent counts are not well explained by the regressors included in the model. According to the results climate change adaptation technologies (Y02A), technologies related to energy generation, transmission or distribution (Y02E), and technologies in the production or processing of goods (Y02P) patent counts vary positively with both the share of highly skilled human capital and the share of foreign inventors. The corresponding IRRs and margins are available in Tables A8 and A10 of the appendix, respectively. With regard to each Y02 subclass, the magnitudes of the predicted margins at the means are of much smaller magnitude compared to GT patents overall. For Y02A, Y02E and Y02P patents, holding all else constant, the predicted counts are less than 60. Interestingly, in the case of Y02P patents, the share of highly skilled foreign workers is negative and statistically significant, indicating that increased shares may decrease patent counts of these technologies. Negative relationships between the share of foreign inventors are not observed in any of the regressions.

Table 7: Negative Binomial Regressions by Y02 Type

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
VARIABLES	Y02A	Y02B	Y02C	Y02D	Y02E	Y02P	Y02T	Y02W
GT	0.00172**	0.0117	0.0217**	0.00205	0.0190**	0.0131**	0.0253***	0.00437
	(0.000876)	(0.00857)	(0.00943)	(0.00416)	(0.00793)	(0.00570)	(0.00584)	(0.00641)
Non-GT	-2.06e-05	-0.000200	-0.000667**	0.000383	-0.000520**	-0.000138	-0.000531***	-4.53e-05
	(7.19e-05)	(0.000268)	(0.000306)	(0.000300)	(0.000251)	(0.000200)	(0.000203)	(0.000242)
High Skilled Foreign = L,	0.871	-4.442	-1.535	3.931	1.004	-4.010**	-3.462	3.133
	(1.578)	(2.986)	(4.453)	(3.091)	(2.008)	(1.878)	(2.271)	(3.035)
Foreign Inventors = L,	4.432**	4.070	-1.111	3.472	4.337***	4.950**	1.573	3.161
	(1.767)	(2.586)	(3.752)	(3.486)	(1.653)	(1.987)	(1.622)	(2.610)
High Skilled HC = L,	13.96***	15.46***	16.02***	28.17***	15.75***	11.04***	6.199*	1.055
	(2.877)	(4.991)	(5.989)	(7.100)	(4.093)	(3.504)	(3.323)	(3.570)
Universities	0.0226***	0.0156***	0.00510	0.000392	0.00529	0.00974	0.00451	0.0141
	(0.00620)	(0.00585)	(0.00381)	(0.00722)	(0.00583)	(0.00622)	(0.00600)	(0.00983)
Population	-6.95e-08	-2.75e-07	2.86e-07	4.12e-07	3.06e-07	6.42e-08	3.38e-07	-1.45e-07
	(2.87e-07)	(2.29e-07)	(7.14e-07)	(3.05e-07)	(3.33e-07)	(2.36e-07)	(4.10e-07)	(2.83e-07)
Constant	-2.070***	-1.818***	-3.599***	-4.155***	-1.810***	-0.805**	-0.649*	-1.483***
	(0.311)	(0.521)	(0.581)	(0.825)	(0.401)	(0.347)	(0.341)	(0.379)
Observations	238	238	238	238	238	238	238	238

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Despite the apparent relationship between patent counts and the share of foreign inventors, Table A11 in the appendix shows that no relationship is present between inventors and the RTAs of certain Y02 subclasses. However, another question arises which is if the shares of these individuals affect the number of separate RTAs by U.S. MSA. Table 8 displays another adaptation of equation (3) in which a negative binomial regression is performed where the dependent variable is a variable of the sum of unique RTAs by MSA. In contrast to the insignificant findings from Table A11 in the appendix, both the shares of highly skilled human capital and foreign inventors are positive and statistically significant. Therefore, holding all other variables constant, the predicted margins from column (2) show that the average predicted number of unique RTAs is about 2.6 for the mean of the share of foreign inventors. Even though neither the shares of foreign inventors nor the share of highly skilled human capital can explain the RTAs in Y02 of MSAs in 2006, they do seem to be positively associated with the number of unique RTAs. This indicates that these shares may help MSAs to specialize in a wider variety of technologies than those with lower shares.

Table 8: Negative Binomial Regression for Cumulative RTA

VARIABLES	(1) Cum. RTA	(2) Margins
GT	0.00166**	0.00238**
	(0.000725)	(0.00106)
Non-GT	-0.000153***	-0.000219***
	(5.78e-05)	(8.40e-05)
High Skilled Foreign = L,	-0.431	-0.616
	(0.874)	(1.255)
Foreign Inventors = L,	1.809**	2.585*
	(0.914)	(1.328)
High Skilled HC = L,	5.023***	7.178***
	(1.308)	(1.872)
Universities	0.000804	0.00115
	(0.00160)	(0.00228)
Population	1.07e-07**	1.54e-07**
	(4.45e-08)	(6.43e-08)
Constant	-0.375**	
	(0.149)	
Observations	238	238

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

5.1 Robustness Checks

A crucial robustness check to this analysis is the consideration of the impact of shares of highly skilled *natives* and share of *native* inventors. Were the reproduction of the regressions with shares of natives to result in significant coefficients and similar signs and magnitudes, then no conclusions could be drawn regarding the particular impact of shares of foreigners. Tables 9-11 serve to this end, reproducing the estimations from Tables 6-8 but using the share of highly skilled natives and the share of native inventors. Consistent with the results from the said tables, the share of highly skilled human capital continues to maintain its relevance and the significance and magnitudes of the other controls do not vary greatly. The shares of natives, however, generally display the opposite results as with foreigners, suggesting that in the cases in which the parameters are significant, increased shares of highly skilled natives and inventors may be negatively associated with patents. This robustness check may also lend its explanation to the theory that foreign skilled workers and inventors bring new, non-redundant into the host societies which may help with the development of green technology as it is more complex and requires a more diversified knowledge base.

Table 9: Negative Binomial Regressions by Patents Type with Native Inventors

	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	Total	Total	GT	GT	Y02	Y02
High Skilled Native = L,	-1.291	0.0639	-0.271	2.050	-0.530	1.208
	(1.196)	(1.225)	(1.461)	(1.557)	(1.265)	(1.275)
Native Inventors = L,		-4.282**		-4.673**		-4.080***
		(1.756)		(1.856)		(1.410)
High Skilled HC = L,	21.38***	18.28***	19.16***	15.04***	15.41***	12.74***
	(2.729)	(2.984)	(3.171)	(3.489)	(2.615)	(2.620)
Universities	0.0244***	0.0239***	0.0234***	0.0243***	0.0101*	0.0101**
	(0.00525)	(0.00526)	(0.00739)	(0.00691)	(0.00517)	(0.00510)
Population	2.15e-08	-2.81e-09	-2.18e-07	-2.15e-07	1.03e-07	1.08e-07
	(2.00e-07)	(2.06e-07)	(2.01e-07)	(2.07e-07)	(2.37e-07)	(2.42e-07)
Non-GT			0.000219	0.000201*	-0.000260	-0.000268
			(0.000140)	(0.000114)	(0.000178)	(0.000164)
GT					0.0150***	0.0147***
					(0.00508)	(0.00509)
Constant	3.291***	6.217***	0.499	3.000	0.675	3.041**
	(1.165)	(1.620)	(1.396)	(1.942)	(1.193)	(1.486)
Observations	238	238	238	238	238	238

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 10: Negative Binomial Regressions by Y02 Type with Native Inventors

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
VARIABLES	Y02A	Y02B	Y02C	Y02D	Y02E	Y02P	Y02T	Y02W
GT	0.00172**	0.0117	0.0217**	0.00205	0.0190**	0.0131**	0.0253***	0.00437
	(0.000876)	(0.00857)	(0.00943)	(0.00416)	(0.00793)	(0.00570)	(0.00584)	(0.00641)
Non-GT	-2.06e-05	-0.000200	-0.000667**	0.000383	-0.000520**	-0.000138	-0.000531***	-4.53e-05
	(7.19e-05)	(0.000268)	(0.000306)	(0.000300)	(0.000251)	(0.000200)	(0.000203)	(0.000242)
High Skilled Native = L,	-0.871	4.442	1.535	-3.931	-1.004	4.010**	3.462	-3.133
	(1.578)	(2.986)	(4.453)	(3.091)	(2.008)	(1.878)	(2.271)	(3.035)
Native Inventors = L,	-4.432**	-4.070	1.111	-3.472	-4.337***	-4.950**	-1.573	-3.161
	(1.767)	(2.586)	(3.752)	(3.486)	(1.653)	(1.987)	(1.622)	(2.610)
High Skilled HC = L,	13.96***	15.46***	16.02***	28.17***	15.75***	11.04***	6.199*	1.055
	(2.877)	(4.991)	(5.989)	(7.100)	(4.093)	(3.504)	(3.323)	(3.570)
Universities	0.0226***	0.0156***	0.00510	0.000392	0.00529	0.00974	0.00451	0.0141
	(0.00620)	(0.00585)	(0.00381)	(0.00722)	(0.00583)	(0.00622)	(0.00600)	(0.00983)
Population	-6.95e-08	-2.75e-07	2.86e-07	4.12e-07	3.06e-07	6.42e-08	3.38e-07	-1.45e-07
	(2.87e-07)	(2.29e-07)	(7.14e-07)	(3.05e-07)	(3.33e-07)	(2.36e-07)	(4.10e-07)	(2.83e-07)
Constant	3.232	-2.189	-6.246	3.248	3.531*	0.135	-2.537	4.810*
	(2.045)	(2.348)	(3.989)	(3.186)	(1.978)	(1.970)	(2.218)	(2.465)
Observations	238	238	238	238	238	238	238	238

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 11: Negative Binomial Regression for Cumulative RTA with Native Inventors

	(1)
VARIABLES	Cum. RTA
GT	0.00166**
	(0.000725)
Non-GT	-0.000153***
	(5.78e-05)
High Skilled Native = L,	0.431
	(0.874)
Native Inventors = L,	-1.809**
	(0.914)
High Skilled HC = L,	5.023***
	(1.308)
Universities	0.000804
	(0.00160)
Population	1.07e-07**
	(4.45e-08)
Constant	1.003
	(0.958)
Observations	238

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

As a final robustness check, this section also considers the possibility of a small numbers problem regarding the shares of foreign inventors and outliers. In certain cases, foreign inventors may be overrepresented in MSAs for which the overall number of inventors is relatively small. For example, an MSA which only has

2 inventors, one which is native and another which is foreign, would have a share of foreign inventors of 0.5. Even when controlling for population, a figure such as this may bias the results in favor of the shares of foreign inventors. To account for this fact, the main estimations from Table 6 are repeated controlling for it. Table 12 presents negative binomial regressions excluding MSAs for which the number of total inventors was in the 10th percentile. The results in columns (2), (4) and (6) display virtually no change in the statistical significance of the foreign inventors parameter and the overall magnitudes drop by less than 1.

Table 12: Negative Binomial Regressions by Patents Type Controlling for Inventor Outliers

	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	Total	Total	GT	GT	Y02	Y02
High Skilled Foreign = L,	1.226	-0.0213	-0.166	-2.288	0.282	-1.299
	(1.167)	(1.227)	(1.395)	(1.554)	(1.200)	(1.237)
Foreign Inventors = L,		4.099**		4.447***		3.879***
		(1.658)		(1.687)		(1.227)
High Skilled HC = L,	20.59***	17.62***	17.61***	13.76***	14.15***	11.65***
	(2.713)	(2.870)	(3.079)	(3.300)	(2.503)	(2.412)
Universities	0.0228***	0.0224***	0.0215***	0.0224***	0.00888*	0.00900*
	(0.00508)	(0.00510)	(0.00708)	(0.00664)	(0.00485)	(0.00479)
Population	3.91e-09	-2.20e-08	-2.16e-07	-2.19e-07	8.35e-08	8.14e-08
	(1.91e-07)	(1.95e-07)	(1.90e-07)	(1.94e-07)	(2.22e-07)	(2.25e-07)
Non-GT			0.000233*	0.000215*	-0.000228	-0.000235
			(0.000136)	(0.000111)	(0.000170)	(0.000160)
GT					0.0142***	0.0139***
					(0.00488)	(0.00491)
Constant	2.202***	2.189***	0.548	0.672*	0.425	0.432*
	(0.251)	(0.240)	(0.347)	(0.344)	(0.263)	(0.247)
Observations	220	220	220	220	220	220

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Alongside the concern of a small numbers problem, this section also considers outliers. There is some debate as to whether outliers in count data should be removed and there appears to be little to no consensus on how to do so. As shown in Figure 1, the total and GT patent count data are right skewed likely as a result of a large number of lower values and zeros in addition to a few unusually high counts. For the purpose of this exercise, z-scores are used to identify those data points which may be considered outlier in lieu of a more concrete method to identify them. Tables 13 and 14 present the repetition of the regressions from Tables 6 and 7 however excluding the MSAs which had GT patent counts outside of three standard deviations from the mean and those which had total counts of inventors in the 10th percentile. Equations (2), (4) and (6) of Table 13 reveal that the significance of the share of foreign inventors remains largely unchanged. However, the magnitudes have dropped a significant amount in comparison to the regressions from Table 6. This is unsurprising considering that some MSA counts of GT patents were close to 1,000 while the mean was around 29. Hence the significantly reduced magnitudes likely reflect the impact these high counts had on the data.

Table 13: Negative Binomial Regressions by Patents Type Controlling for GT Outliers

	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	Total	Total	GT	GT	Y02	Y02
High Skilled Foreign = L,	-0.523	-1.901	-0.138	-2.130	0.982	-0.375
	(1.474)	(1.446)	(1.363)	(1.511)	(1.097)	(1.217)
Foreign Inventors = L,		3.941**		3.849**		3.039**
		(1.542)		(1.700)		(1.185)
High Skilled HC = L,	18.88***	16.04***	16.46***	13.34***	11.22***	9.508***
	(2.501)	(2.731)	(2.947)	(3.173)	(1.980)	(2.014)
Universities	0.0282***	0.0276***	0.0202***	0.0212***	0.00916**	0.00978***
	(0.00435)	(0.00434)	(0.00523)	(0.00516)	(0.00359)	(0.00348)
Population	2.11e-07	1.95e-07	-8.74e-08	-6.06e-08	6.81e-08	9.07e-08
	(1.88e-07)	(1.91e-07)	(1.89e-07)	(1.95e-07)	(1.96e-07)	(2.02e-07)
Non-GT			0.000424***	0.000383***	-0.000237	-0.000283**
			(0.000122)	(0.000113)	(0.000174)	(0.000137)
GT					0.0206***	0.0204***
					(0.00321)	(0.00307)
Constant	2.376***	2.383***	0.531	0.630*	0.414*	0.409*
	(0.285)	(0.268)	(0.334)	(0.334)	(0.237)	(0.228)
Observations	214	214	214	214	214	214

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 14: Negative Binomial Regressions by Y02 Type Controlling for GT Outliers

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
VARIABLES	Y02A	Y02B	Y02C	Y02D	Y02E	Y02P	Y02T	Y02W
GT	0.00149	0.0232***	0.0412***	0.00957	0.0279***	0.0180***	0.0324***	0.00991
	(0.00309)	(0.00699)	(0.0104)	(0.00868)	(0.00558)	(0.00464)	(0.00467)	(0.00604)
Non-GT	2.37e-05	-0.000577**	-0.00180***	0.000567	-0.000779***	-6.75e-05	-0.000622***	-6.68e-05
	(0.000149)	(0.000262)	(0.000472)	(0.000400)	(0.000225)	(0.000261)	(0.000205)	(0.000217)
High Skilled Foreign = L,	1.395	-3.210	-0.107	4.871*	1.528	-2.800	-2.265	4.145
	(1.738)	(3.047)	(4.880)	(2.856)	(1.811)	(1.858)	(2.398)	(3.037)
Foreign Inventors = L,	4.106**	2.297	-1.838	0.792	4.291***	3.791*	0.543	3.207
	(1.600)	(2.494)	(4.237)	(3.412)	(1.499)	(1.949)	(1.672)	(2.494)
High Skilled HC = L,	11.88***	11.11***	12.87**	23.52***	11.90***	7.850**	3.485	-1.581
	(2.656)	(3.904)	(5.862)	(6.665)	(3.083)	(3.069)	(3.224)	(3.638)
Universities	0.0265***	0.0186***	0.00756	-0.00484	0.00573	0.00797*	0.00409	0.0120
	(0.00611)	(0.00502)	(0.00894)	(0.00690)	(0.00468)	(0.00471)	(0.00506)	(0.00761)
Population	-2.98e-08	-3.92e-07*	8.54e-07	5.01e-07	3.85e-07	8.77e-08	2.57e-07	-3.10e-07
	(2.98e-07)	(2.25e-07)	(7.23e-07)	(3.85e-07)	(2.89e-07)	(2.27e-07)	(3.88e-07)	(3.59e-07)
Constant	-1.932***	-1.531***	-3.953***	-3.717***	-1.640***	-0.623*	-0.493	-1.366***
	(0.309)	(0.462)	(0.622)	(0.788)	(0.366)	(0.327)	(0.339)	(0.379)
Observations	214	214	214	214	214	214	214	214

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 14 also appears to mostly preserve the results reported in Table 7, however, the share of foreign inventors is no longer significant for Y02P patents. This suggests that the observations which were excluded in this exercise contained a large number of Y02P patents. Intuitively, those MSAs with higher counts of GT patents would be expected to also have higher counts of Y02 patents since the two variables share similar means and standard deviations. Despite the practice of excluding outliers from other forms of estimation, in the case of patent counts it may be dangerous. Among the observations excluded for unusually high GT patent counts are MSAs which encompass key innovation hubs such as Chicago, IL, Silicon Valley, Boston, MA and Detroit, MI. While they may be outliers, statistically speaking, they also represent innovation/patenting hubs. Removing them from the analysis may have some justification in statistical terms, however, intuitively, it may not make sense to remove them since they may be the areas responsible for a considerable share of U.S. innovation.

6 Conclusion

Climate change continues to threaten the health of the planet and its inhabitants. The development and implementation of a wide variety of new technologies which aim to mitigate climate change and its associated causes will be essential for reducing the damage. Green technology will therefore play a central role in this fight. Understanding which sort of innovation currently constitutes the base of such technologies and understanding how they differ is crucial for determining which factors influence their development. Previous literature has suggested that migration and the presence of foreign inventors is responsible for augmenting shares of crucial types of technology and the development of new and more complex ones. According to the previous literature green technologies are often more novel and more complex as they combine a greater variety of existing knowledge. Therefore, understanding these challenges in the context of GT is particularly important.

The foregoing analysis demonstrates that in the context of U.S. MSAs, there are certain characteristics which are associated with the development of green technologies. It has been clearly shown that the shares of skilled and highly skilled foreign workers and the share of foreign inventors is among these factors. In nearly all cases, the share of foreign inventors is associated with increased counts of GT patents in posterior periods. Additionally, there is evidence that this effect is specific to certain types of GT and may even promote further diversification into more forms of GT.

Despite these important findings, this study presents certain limitations. Firstly, the unavailability of data to construct a panel dataset presents a number of concerns. One is that more recent years may be more suitable for this analysis. Another is the utility of panel data to apply fixed effects to control for unobserved MSA heterogeneity. A second limitation is the issue of the potential endogeneity of the shares of foreign inventors and human capital. Migration may be a decision and certain locations may be more likely to attract certain types of individuals based on a number of factors. In turn, this may lead to correlation between residuals and the error term which bias results. Ideally, an instrumental variable approach would be used to correct for this endogeneity. However, given the scope of this thesis and the availability of data, none was able to be successfully applied.

While this thesis provides proximate evidence for the influence of the diverse knowledge sets that foreign and migrant inventors import on their local innovator setting, a better understanding of the mechanisms which drive this change is left unstudied. Future research must therefore be conducted which not only investigates the relationships of these variables, but also seeks to explain the mechanisms behind them. For example, is there a certain place of origin of foreign inventors or a certain knowledge base from their home country which is promoting diversification into specific types of green technologies? An understanding of these complex pathways may lead to the development of more targeted immigration policy to attract experts in certain fields crucial to the development of technologies which are most crucial to the fight against climate change.

References

Aghion, P., Hemous, D., & Veugelers, R. (2009). No Green Growth without Innovation. Bruegel Policy Brief, 2009(7).

Alesina, A., Harnoss, J., & Rapoport, H. (2016). Birthplace diversity and economic prosperity. *Journal of Economic Growth*, 21(2), 101-138.

Angelucci, S., Hurtado-Albir, F. J., & Volpe, A. (2018). Supporting global initiatives on climate change: The EPO's "Y02-Y04S" tagging scheme. *World Patent Information*, *54*, S85-S92.

Bahar, D., Rapoport, H., & Turati, R. (2019). Does birthplace diversity affect economic complexity? Cross-country evidence. Cross-country evidence.

Barbieri, N., Marzucchi, A., Rizzo, U. (2020). Knowledge sources and impacts on subsequent inventions: do green technologies differ from non-green ones? *Res. Policy*, 49 (2020), p.103901, 10.1016/j.respol.2019.103901

Bathelt, H., Malmberg, A., & Maskell, P. (2004). Clusters and knowledge: local buzz, global pipelines and the process of knowledge creation. *Progress in human geography*, 28(1), 31-56.

Chellaraj, G., Maskus, K.,. Mattoo, A. (2005). The Contribution of Skilled Immigration and International Graduate Students to U.S. Innovation. Policy Research Working Paper; No. 3588. World Bank, Washington, DC. © World Bank. https://openknowledge.worldbank.org/handle/10986/8957 License: CC BY 3.0 IGO.

Chen, A., & Chen, H. (2021). Decomposition Analysis of Green Technology Innovation from Green Patents in China. *Mathematical Problems in Engineering*, *2021*, 1-11.

Desheng, L., Jiakui, C., & Ning, Z. (2021). Political connections and green technology innovations under an environmental regulation. *Journal of Cleaner Production*, *298*, 126778.

Durán-Romero, G., & Urraca-Ruiz, A. (2015). Climate change and eco-innovation. A patent data assessment of environmentally sound technologies. *Innovation*, *17*(1), 115-138.

Favot, M., Vesnic, L., Priore, R., Bincoletto, A., & Morea, F. (2023). Green patents and green codes: How different methodologies lead to different results. *Resources, Conservation & Recycling Advances*, *18*, 200132.

Ganguli, I. (2015). Immigration and ideas: what did Russian scientists "bring" to the United States?. *Journal of Labor Economics*, 33(S1), S257-S288.

Haines, A., & Patz, J. A. (2004). Health effects of climate change. *JAMA*. 291(1), 99–103. https://doi.org/10.1001/jama.291.1.99

Hall, B. & Helmers, C. (2010). The role of patent protection in (clean/green) technology transfer. No 2010-046, MERIT Working Papers, United Nations University – Maastricht Economic and Social Research Institute on Innovation and Technology (MERIT)

Haščič, I., & Migotto, M. (2015). Measuring environmental innovation using patent data. OECD.

Hornung, E. (2014). Immigration and the diffusion of technology: The Huguenot diaspora in Prussia. *American Economic Review*, 104(1), 84-122.

Hunt, J., & Gauthier-Loiselle, M. (2010). How Much Does Immigration Boost Innovation? *American Economic Journal: Macroeconomics*, 2 (2): 31-56. DOI:10.1257/mac.2.2.31

Kemeny, T. (2017). Immigrant diversity and economic performance in cities. *International Regional Science Review*, *40*(2), 164-208.

Kraus, S., Rehman, S. U., & García, F. J. S. (2020). Corporate social responsibility and environmental performance: The mediating role of environmental strategy and green innovation. *Technological Forecasting and Social Change*, *160*, 120262.

Lissoni, Francesco, (2018), International migration and innovation diffusion: an eclectic survey, *Regional Studies*, 52(5), p. 702-714,

Maraut, S., Dernis, H., Webb, C., Spiezia, V., Guellec, D., (2008). The OECD REGPAT Database (OECD Science, Technology and Industry WorkingPapers). Organisation for Economic Co-operation and Development, Paris.

McMichael, A. J., Woodruff, R. E., & Hales, S. (2006). Climate change and human health: present and future risks. *Lancet* (London, England), 367(9513), 859–869. https://doi.org/10.1016/S01406736(06)68079-3
Miguelez, E., & Morrison, A. (2022). Migrant inventors as agents of technological change. *The Journal of Technology Transfer*, 1-24.
Miguelez, E., & Fink, C. (2013). *Measuring the international mobility of inventors: A new database* (Vol. 8). WIPO.
Montresor, S. & Quatraro, F. (2020). Green technologies and Smart Specialisation Strategies: a European patent-based analysis of the intertwining of technological relatedness and key enabling technologies, Regional Studies, 54:10, 1354-1365, DOI: 10.1080/00343404.2019.1648784
Morrison, A., Rabellotti, R., & Zirulia, L. (2013). When do global pipelines enhance the diffusion of knowledge in clusters?. *Economic geography*, *89*(1), 77-96.
Moser, P., Voena, A., & Waldinger, F. (2014). German Jewish émigrés and US invention. *American Economic Review*, 104(10), 3222-55.
National Research Council. (2012). *Research universities and the future of America: Ten breakthrough actions vital to our nation's prosperity and security*. National Academies Press.
OECD, 2022. ENV-TECH - Patent search Strategies For the Identification of Selected Environment-Related Technologies. Ivan Haščič and Mauro Migotto.
Orsatti, G., Quatraro, F., & Pezzoni, M. (2020). The antecedents of green technologies: The role of team-level recombinant capabilities. *Research Policy*, 49(3), 103919.
Ozgen, C., Nijkamp, P., & Poot, J. (2012). Immigration and innovation in European regions. In Migration impact assessment. Edward Elgar Publishing.
Perruchas, F., Consoli, D., & Barbieri, N. (2020). Specialisation, diversification and the ladder of green technology development. *Research Policy*, *49*(3), 103922.
Queen’s University. (n.d.). *Research guides: Patents and designs: Classification systems*. Classification Systems - Patents and Designs - Research Guides at Queen’s University Library. https://guides.library.queensu.ca/patents/
Stephan, P., Levin, S.G. (2001). Exceptional contributions to US science by the foreign-born and foreign-educated. *Population Research and Policy Review* 20, 59–79. https://doi.org/10.1023/A:1010682017950
Steven Ruggles, Sarah Flood, Matthew Sobek, Danika Brockman, Grace Cooper, Stephanie Richards, and Megan Schouweiler. IPUMS USA: Version 13.0 [dataset]. Minneapolis, MN: IPUMS, 2023. https://doi.org/10.18128/D010.V13.0
Intergovernmental Panel on Climate Change. (2022). The Evidence Is Clear: The Time for Action Is Now. We Can Halve Emissions by 2030. https://www.ipcc.ch/2022/04/04/ipcc-ar6-wgiii-pressrelease/
U.S. Census Bureau. (2022). *Tigerlines/Shapefiles.* Retrieved May 3, 2023, from https://www.census.gov/geographies/mapping-files/time-series/geo/tiger-line-file.html
Zeppini, P., & van den Bergh, J. (2011). Competing Recombinant Technologies for Environmental Innovation: Extending Arthur's Model of Lock-In. *Industry and Innovation*, *18*(3), 317-334. https://doi.org/10.1080/13662716.2011.561031

Appendix

Table A1: Matrix of correlations

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)
(1) Total Patents	1.00																
(2) GT	0.78	1.00															
(3) Y02	0.84	0.99	1.00														
(4) Y02A	0.71	0.66	0.75	1.00													
(5) Y02B	0.84	0.69	0.75	0.67	1.00												
(6) Y02C	0.34	0.41	0.41	0.31	0.39	1.00											
(7) Y02D	0.88	0.60	0.66	0.47	0.72	0.17	1.00										
(8) Y02E	0.75	0.87	0.88	0.66	0.76	0.42	0.53	1.00									
(9) Y02P	0.85	0.77	0.82	0.67	0.70	0.42	0.67	0.73	1.00								
(10) Y02T	0.25	0.75	0.67	0.28	0.19	0.19	0.12	0.50	0.22	1.00							
(11) Y02W	0.40	0.49	0.50	0.46	0.38	0.23	0.24	0.44	0.47	0.23	1.00						
(12) Lagged Share of Skilled Foreigners	0.54	0.41	0.45	0.42	0.43	0.16	0.49	0.41	0.45	0.11	0.28	1.00					
(13) Lagged Share of High Skilled Foreigners	0.53	0.42	0.46	0.39	0.42	0.17	0.49	0.41	0.45	0.13	0.29	0.93	1.00				
(14) Lagged Share of Foreign Inventors	0.35	0.31	0.34	0.29	0.31	0.14	0.33	0.35	0.32	0.11	0.17	0.50	0.57	1.00			
(15) Lagged Share of High Skilled Human Capital	0.46	0.40	0.44	0.42	0.42	0.26	0.39	0.45	0.40	0.14	0.18	0.35	0.40	0.47	1.00		
(16) Universities	0.63	0.63	0.66	0.71	0.68	0.44	0.41	0.60	0.62	0.25	0.57	0.43	0.40	0.24	0.31	1.00	
(17) Population	0.38	0.39	0.42	0.51	0.43	0.17	0.28	0.42	0.34	0.14	0.65	0.42	0.35	0.19	0.13	0.62	1.00

Table A2: OLS Regressions – Total Patents

	(1)	(2)	(3)	(4)	(5)
VARIABLES	Total	Total	Total	Total	Total
Skilled Foreign = L,	5,549***	3,456*			
	(2,132)	(1,836)			
High Skilled Foreign = L,	2,298	746.0	3,605***	3,605***	3,767***
	(1,966)	(1,673)	(707.0)	(707.0)	(797.9)
High Skilled HC = L,		5,567***	5,368***	5,368***	5,562***
		(1,314)	(1,317)	(1,317)	(1,390)
Universities		10.83***	11.01***	11.01***	10.99***
		(1.401)	(1.406)	(1.406)	(1.409)
Population		-8.10e-05	-5.42e-05	-5.42e-05	-5.41e-05
		(7.44e-05)	(7.34e-05)	(7.34e-05)	(7.35e-05)
Foreign Inventors = L,					-313.7
					(710.6)
Constant	-465.7***	-856.6***	-847.9***	-847.9***	-852.4***
	(104.4)	(131.6)	(132.2)	(132.2)	(132.8)
Observations	238	238	238	238	238
R-squared	0.299	0.529	0.522	0.522	0.522

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table A3: OLS Regressions – GT Patents

	(1)	(2)	(3)	(4)	(5)
VARIABLES	Total	Total	Total	Total	Total
Non-GT	0.0418***	0.0337***	0.0330***	0.0330***	0.0331***
	(0.00267)	(0.00314)	(0.00313)	(0.00313)	(0.00314)
Skilled Foreign = L,	-210.0	-314.9*			
	(168.5)	(167.4)			
High Skilled Foreign = L,	220.0	246.5	-9.630	-9.630	-41.31
	(153.9)	(151.7)	(67.22)	(67.22)	(75.28)
High Skilled HC = L,		142.6	166.9	166.9	128.9
		(122.9)	(122.9)	(122.9)	(129.4)
Universities		0.505***	0.504***	0.504***	0.506***
		(0.144)	(0.145)	(0.145)	(0.145)
Population		5.89e-06	3.34e-06	3.34e-06	3.35e-06
		(6.77e-06)	(6.67e-06)	(6.67e-06)	(6.67e-06)
Foreign Inventors = L,					60.19
					(64.33)
Constant	9.646	-6.060	-7.861	-7.861	-6.875
	(8.474)	(12.87)	(12.91)	(12.91)	(12.95)
Observations	238	238	238	238	238
R-squared	0.599	0.637	0.632	0.632	0.633

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table A2 and A3 list the sequential OLS regression of equations (1) and (2), first implementing the share of skilled foreign workers, then including the share of highly skilled foreign workers and eventually the share of foreign inventors. The interpretation of the results of this regression are largely irrelevant as an OLS estimator in not appropriate in this context. However, the significance of some of the variables indicates which should be included in the other estimations and this regression may be used as well to address any concerns of multicollinearity which may bias the results. Table A4 lists the variable inflation factors (VIF) for each of the primary regressors. Notably, none of the variables present a VIF which are of concern (greater than five) so we can conclude that multicollinearity is not of concern in this case.

Table A4: Variable Inflation Factor (VIF)

VARIABLES	VIF	1/VIF
Non-GT	2.070	0.484
High Skilled Foreign	1.820	0.548
Universities	1.800	0.556
Foreign Inventors	1.760	0.569
High Skilled HC	1.570	0.638
Population	1.260	0.793
Mean VIF	1.710	

Another potential concern in the following regressions is that of heteroskedasticity of the residuals. Certain MSA observable characteristics may be responsible for variations in patent counts. Even though these unobserved characteristics cannot be controlled for in the case of this cross-section, the potential effect that the resulting heteroskedasticity has on the calculation of the standard errors in the regressions should be accounted for. Figure A1 displays the graph of the predicted residuals from estimation of equation (1) in equation (4) from Table A2. While the residuals do not follow a particular trend, they are clearly not evenly distributed and therefore display some level of heteroskedasticity.

Figure A1: OLS Estimation Residuals

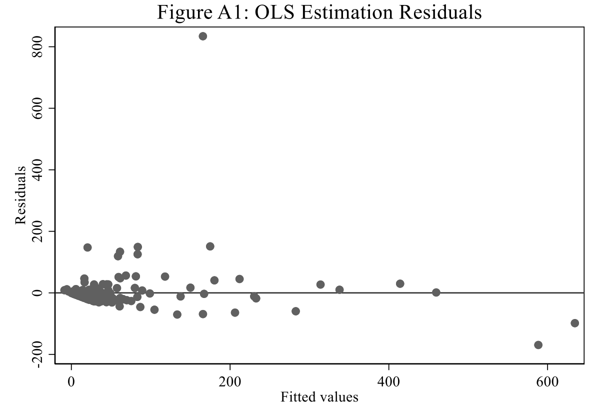
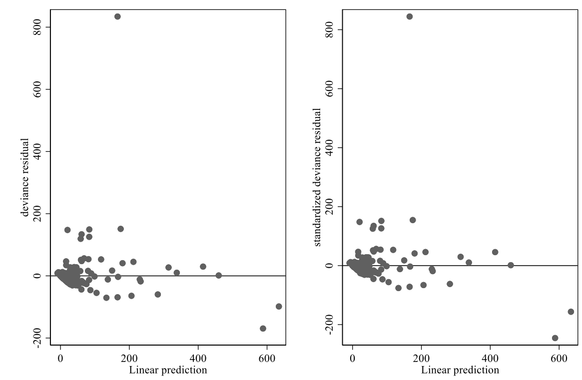


Figure A2: GLM Estimation Residuals



Accordingly, the distribution of the residuals from a generalized linear model (GLM) should be checked to see how they compare as this type of estimation is more similar to the Poisson and negative binomial estimation methods. Figure A2 presents the plots of these residuals using both the actual and predicted values of GT patent counts. The graphs display similar distributions to the residuals from the OLS estimation and again indicate the presence of heteroskedasticity. Even though count data usually does not follow a normal distribution and constant variance is not an assumption of the Poisson or negative binomial models, its potential influence on the calculation of the standard errors should be taken into account. Therefore, the subsequent estimations in this paper will be conducted correcting for this heteroskedasticity using clustered standard errors at the MSA level.

This thesis made a case based on various statistics and tests to estimate the focal equations by a negative binomial regression instead of a Poisson regression. The overdispersion of the dependent variables indicated that we should favor a negative binomial estimation. Tables A5 and A6 serve as a justification for the methodology and a robustness check for the results. Comparing the reported coefficients, the statistical significance of the explanatory regressors and their corresponding magnitudes do not vary greatly. In fact, some statistical significance of the share of foreign inventors parameter may have been lost in favor of a model which more appropriately estimates the equations. Notably, the share of foreign inventors is no longer statistically significant in column (2) of Table A5 where the dependent variable is the total patent count. Differences in the statistical significance of the parameters are also observed in Table

A6 with regard to the counts of different Y02 technologies. Despite some slightly different results, the Poisson regressions serve to show that the findings under the negative binomial estimations still hold.

Table A5: Poisson Regressions by Patents Type with Inventors

	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	Total	Total	GT	GT	Y02	Y02
High Skilled Foreign = L,	5.351***	4.294***	-0.872	-3.185*	-1.795	-4.415***
	(0.785)	(1.027)	(1.655)	(1.843)	(1.418)	(1.625)
Foreign Inventors = L,		2.556*		4.245***		4.913***
		(1.353)		(1.022)		(1.031)
High Skilled HC = L,	12.67***	11.76***	9.870***	7.932***	12.00***	9.714***
	(2.118)	(2.151)	(2.344)	(2.251)	(1.435)	(1.419)
Universities	0.00518***	0.00535***	0.00348**	0.00349**	0.00223**	0.00227**
	(0.000987)	(0.00102)	(0.00153)	(0.00152)	(0.00108)	(0.00107)
Population	5.90e-08	6.04e-08	1.24e-07*	1.32e-07*	1.16e-07**	1.27e-07***
	(4.67e-08)	(4.54e-08)	(6.94e-08)	(6.85e-08)	(4.66e-08)	(4.47e-08)
Non-GT			0.000142***	0.000160***	5.21e-05*	7.41e-05**
			(3.42e-05)	(3.47e-05)	(3.01e-05)	(3.17e-05)
GT					0.00372***	0.00368***
					(0.000261)	(0.000245)
Constant	3.204***	3.099***	2.239***	2.180***	1.774***	1.691***
	(0.230)	(0.232)	(0.278)	(0.285)	(0.191)	(0.201)
Observations	238	238	238	238	238	238

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table A6: Poisson Regressions by Y02 Type

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
VARIABLES	Y02A	Y02B	Y02C	Y02D	Y02E	Y02P	Y02T	Y02W
GT	0.00258***	0.00196***	0.00304***	0.00121**	0.00348***	0.00285***	0.00561***	0.00226***
	(0.000261)	(0.000628)	(0.000589)	(0.000606)	(0.000360)	(0.000658)	(0.000228)	(0.000737)
Non-GT	6.07e-05	0.000164***	-4.06e-05	0.000225***	5.96e-05	0.000121	-1.75e-05	-1.47e-05
	(5.06e-05)	(4.25e-05)	(9.29e-05)	(6.53e-05)	(3.88e-05)	(7.83e-05)	(8.22e-05)	(8.44e-05)
High Skilled Foreign = L,	-4.723**	-6.431***	-4.265	-5.702*	-5.147**	-4.196	-7.214***	0.788
	(2.220)	(1.931)	(4.741)	(3.153)	(2.123)	(4.202)	(1.689)	(2.448)
Foreign Inventors = L,	5.177**	5.255***	1.374	7.335***	5.682***	5.142***	2.826**	1.166
	(2.580)	(1.632)	(2.602)	(1.933)	(1.980)	(1.640)	(1.438)	(1.366)
High Skilled HC = L,	15.14***	9.911***	15.06***	8.425***	11.76***	7.537***	12.22***	3.214
	(3.073)	(2.397)	(3.454)	(2.065)	(2.145)	(2.625)	(2.974)	(2.467)
Universities	0.00415**	0.00286***	0.00638***	0.00127	0.00195*	0.00283**	0.00107	0.00337
	(0.00164)	(0.000983)	(0.00161)	(0.00121)	(0.00103)	(0.00140)	(0.00221)	(0.00266)
Population	1.81e-07***	1.60e-07***	2.61e-08	1.88e-07***	1.55e-07***	8.88e-08	1.99e-07***	1.72e-07**
	(5.40e-08)	(4.84e-08)	(9.45e-08)	(6.77e-08)	(3.95e-08)	(9.14e-08)	(6.59e-08)	(8.65e-08)
Constant	-0.982**	-0.462*	-2.544***	-0.481	0.107	0.524**	0.359	-0.764**
	(0.396)	(0.274)	(0.479)	(0.382)	(0.321)	(0.263)	(0.458)	(0.358)
Observations	238	238	238	238	238	238	238	238

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table A7: Negative Binomial Regressions IRR by Patent Type

	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	Total	Total	GT	GT	Y02	Y02
High Skilled Foreign = L,	3.637	0.938	1.311	0.129	1.699	0.299
	(4.350)	(1.149)	(1.915)	(0.200)	(2.149)	(0.381)
Foreign Inventors = L,		72.38**		107.1**		59.15***
		(127.1)		(198.7)		(83.41)
High Skilled HC = L,	1.924e+09***	8.690e+07***	2.096e+08***	3.391e+06***	4.947e+06***	340,639***
	(5.250e+09)	(2.593e+08)	(6.647e+08)	(1.183e+07)	(1.294e+07)	(892,476)
Universities	1.025***	1.024***	1.024***	1.025***	1.010*	1.010**
	(0.00538)	(0.00539)	(0.00757)	(0.00708)	(0.00522)	(0.00515)
Population	1.000	1.000	1.000	1.000	1.000	1.000
	(2.00e-07)	(2.06e-07)	(2.01e-07)	(2.07e-07)	(2.37e-07)	(2.42e-07)
GT					1.015***	1.015***
					(0.00515)	(0.00517)
Non-GT			1.000	1.000*	1.000	1.000
			(0.000140)	(0.000114)	(0.000178)	(0.000164)
Constant	7.388***	7.380***	1.256	1.458	1.156	1.184
	(1.833)	(1.769)	(0.436)	(0.502)	(0.319)	(0.304)
Observations	238	238	238	238	238	238

Robust seeform in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table A8: Negative Binomial Regression IRR by Y02 Patent Type

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
VARIABLES	Y02A	Y02B	Y02C	Y02D	Y02E	Y02P	Y02T	Y02W
GT	1.002**	1.012	1.022**	1.002	1.019**	1.013**	1.026***	1.004
	(0.000877)	(0.00867)	(0.00964)	(0.00417)	(0.00809)	(0.00578)	(0.00599)	(0.00644)
Non-GT	1.000	1.000	0.999**	1.000	0.999**	1.000	0.999***	1.000
	(7.19e-05)	(0.000268)	(0.000306)	(0.000300)	(0.000251)	(0.000200)	(0.000203)	(0.000242)
High Skilled Foreign = L,	2.388	0.0118	0.215	50.97	2.729	0.0181**	0.0314	22.94
	(3.770)	(0.0352)	(0.959)	(157.6)	(5.481)	(0.0341)	(0.0712)	(69.62)
Foreign Inventors = L,	84.08**	58.58	0.329	32.21	76.51***	141.2**	4.821	23.59
	(148.6)	(151.5)	(1.235)	(112.3)	(126.5)	(280.6)	(7.819)	(61.55)
High Skilled HC = L,	1.160e+06***	5.184e+06***	9.070e+06***	1.722e+12***	6.907e+06***	62,255***	492.1*	2.873
	(3.338e+06)	(2.588e+07)	(5.432e+07)	(1.223e+13)	(2.827e+07)	(218,126)	(1,635)	(10.26)
Universities	1.023***	1.016***	1.005	1.000	1.005	1.010	1.005	1.014
	(0.00634)	(0.00594)	(0.00383)	(0.00722)	(0.00586)	(0.00628)	(0.00602)	(0.00997)
Population	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	(2.87e-07)	(2.29e-07)	(7.14e-07)	(3.05e-07)	(3.33e-07)	(2.36e-07)	(4.10e-07)	(2.83e-07)
Constant	0.126***	0.162***	0.0273***	0.0157***	0.164***	0.447**	0.523*	0.227***
	(0.0393)	(0.0846)	(0.0159)	(0.0129)	(0.0655)	(0.155)	(0.178)	(0.0860)
Observations	238	238	238	238	238	238	238	238

Robust seeform in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table A9: Negative Binomial Regressions Margins by Patent Type

	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	Total	Total	GT	GT	Y02	Y02
High Skilled Foreign = L,	2,006***	1,610***	-37.70	-137.7*	-73.54	-180.8***
	(344.4)	(414.0)	(69.27)	(72.55)	(57.29)	(65.23)
Foreign Inventors = L,		958.2*		183.6***		201.2***
		(522.5)		(46.56)		(43.13)
High Skilled HC = L,	4,749***	4,408***	426.9***	343.0***	491.5***	397.9***
	(1,027)	(1,007)	(85.39)	(80.25)	(59.22)	(55.49)
Universities	1.941***	2.006***	0.151*	0.151**	0.0916**	0.0929**
	(0.451)	(0.464)	(0.0771)	(0.0769)	(0.0464)	(0.0463)
Population	2.21e-05	2.26e-05	5.37e-06**	5.71e-06**	4.74e-06**	5.18e-06***
	(1.75e-05)	(1.69e-05)	(2.57e-06)	(2.47e-06)	(1.85e-06)	(1.72e-06)
Non-GT			0.00614***	0.00691***	0.00213*	0.00304**
			(0.00142)	(0.00143)	(0.00122)	(0.00129)
GT					0.152***	0.151***
					(0.0154)	(0.0140)
Observations	238	238	238	238	238	238

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table A10: Negative Binomial Regressions Margins by Y02 Type

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
VARIABLES	Y02A	Y02B	Y02C	Y02D	Y02E	Y02P	Y02T	Y02W
GT	0.0146***	0.00733***	0.00166***	0.00644*	0.0346***	0.0292***	0.0499***	0.00314***
	(0.00189)	(0.00248)	(0.000474)	(0.00342)	(0.00529)	(0.00863)	(0.00485)	(0.00113)
Non-GT	0.000344	0.000613***	-2.22e-05	0.00120***	0.000593	0.00124*	-0.000156	-2.04e-05
	(0.000303)	(0.000158)	(5.20e-05)	(0.000384)	(0.000391)	(0.000741)	(0.000729)	(0.000117)
High Skilled Foreign = L,	-26.81*	-24.05***	-2.330	-30.48*	-51.23**	-43.02	-64.17***	1.093
	(14.07)	(7.711)	(2.745)	(17.64)	(23.06)	(40.53)	(16.74)	(3.417)
Foreign Inventors = L,	29.38*	19.65***	0.750	39.20***	56.55***	52.72***	25.14**	1.617
	(16.13)	(6.787)	(1.481)	(12.17)	(21.49)	(15.08)	(11.97)	(1.850)
High Skilled HC = L,	85.95***	37.06***	8.227***	45.03***	117.0***	77.27***	108.7***	4.457
	(21.27)	(9.434)	(2.535)	(9.672)	(25.38)	(22.66)	(27.84)	(3.226)
Universities	0.0236**	0.0107***	0.00349***	0.00677	0.0194*	0.0291*	0.00954	0.00467
	(0.0104)	(0.00405)	(0.00123)	(0.00646)	(0.0104)	(0.0161)	(0.0196)	(0.00377)
Population	1.03e-06***	6.00e-07***	1.42e-08	1.00e-06***	1.55e-06***	9.10e-07	1.77e-06***	2.39e-07**
	(3.07e-07)	(1.76e-07)	(5.18e-08)	(3.62e-07)	(4.05e-07)	(8.79e-07)	(5.97e-07)	(1.13e-07)
Observations	238	238	238	238	238	238	238	238

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table A11 extends the scope of the negative binomial regressions by estimating a logistic regression of a modified version of equation (3) which takes the revealed technological advantage (RTA) of each of the Y02 patent categories as dependent variables. Overall, the results are largely insignificant. Only in the case of the RTA in Y02C technologies is the share of highly skilled human capital positive and statistically significant, indicating that higher shares increase the likelihood that a given MSA develops a RTA in Y02C technologies. This finding may indicate that, while there is some evidence that overall and specific Y02 patent counts increase with the share of foreign inventors and the share of highly skilled human capital, this relationship is not strong enough to provoke strong specialization in these categories.

Table A11: Logistic Regressions by Y02 RTA

VARIABLES	(1) RTA Y02A	(2) RTA Y02B	(3) RTA Y02C	(4) RTA Y02D	(5) RTA Y02E	(6) RTA Y02P	(7) RTA Y02T	(8) RTA Y02W
GT	-0.0107*	0.00221	0.00328	-0.00179	0.0157**	0.00505	0.0407***	0.00126
	(0.00621)	(0.00229)	(0.00228)	(0.00247)	(0.00735)	(0.00723)	(0.0131)	(0.00191)
Non-GT	-0.000250	-0.000306	-0.000486	0.000279	-0.00118*	-0.000386	-0.00124	-0.000241
	(0.000412)	(0.000233)	(0.000346)	(0.000307)	(0.000648)	(0.000339)	(0.000924)	(0.000227)
High Skilled Foreign = L,	2.414	-7.985**	-5.277	6.194	1.853	-4.350	-4.313	0.243
	(3.293)	(3.282)	(6.134)	(4.072)	(2.939)	(3.557)	(4.118)	(2.607)
Foreign Inventors = L,	3.824	3.685	-3.975	-0.435	2.497	3.779*	0.290	3.495
	(2.765)	(2.489)	(4.574)	(2.735)	(2.508)	(2.157)	(3.262)	(2.306)
High Skilled HC = L,	7.817*	9.322*	24.11***	9.676*	7.793	5.723	-3.227	1.754
	(4.721)	(4.850)	(7.002)	(4.996)	(5.445)	(4.995)	(6.687)	(4.819)
Universities	0.0121	0.00941	0.0154	-0.00828	-0.00271	-0.00154	-0.0595**	0.00409
	(0.0108)	(0.00662)	(0.00960)	(0.00904)	(0.0105)	(0.00782)	(0.0233)	(0.00611)
Population	4.04e-07	2.54e-07	-5.81e-08	2.29e-07	1.74e-07	2.54e-07	7.10e-07	3.11e-07
	(2.66e-07)	(2.33e-07)	(2.16e-07)	(2.70e-07)	(3.06e-07)	(2.75e-07)	(5.59e-07)	(2.17e-07)
Constant	-2.750***	-1.994***	-4.157***	-3.890***	-2.601***	-2.013***	-0.991	-2.016***
	(0.513)	(0.437)	(0.640)	(0.535)	(0.577)	(0.503)	(0.662)	(0.479)
Observations	238	238	238	238	238	238	238	238

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1