8-2012

International Knowledge Flows and Technological Advance: The Role of International Migration

Kacey N. Douglas
University of Nebraska-Lincoln

INTERNATIONAL KNOWLEDGE FLOWS

AND TECHNOLOGICAL ADVANCE:

THE ROLE OF INTERNATIONAL MIGRATION

By

Kacey Nicole Douglas

A DISSERTATION

Presented to the Faculty of

The Graduate College at the University of Nebraska

In Partial Fulfillment of Requirements

For the Degree of Doctor of Philosophy

Major: Economics

Under the Supervision of Professor Hendrik van den Berg

Lincoln, Nebraska

August, 2012

INTERNATIONAL KNOWLEDGE FLOWS

AND TECHNOLOGICAL ADVANCE:

THE ROLE OF INTERNATIONAL MIGRATION

Kacey Nicole Douglas, Ph.D.

University of Nebraska, 2012

Advisor: Hendrik van den Berg

Immigration is a major aspect of globalization. As the world becomes increasingly integrated, it becomes important to learn more about the effects of immigration on global economic growth. According to Robert Solow's long run growth model, technological advance is the only form of economic growth sustainable in the long run. Those who contribute to technological advance – highly skilled labor – however, increasingly emigrate from lesser developed to more developed countries in a process known as brain drain. This process has been shown to lead to a permanent increase in income and growth in the host country relative to the source country. This paper investigates whether brain drain migration can lead to technological advance in the source country. More specifically, do migration flows to the United States (US) lead to knowledge from the US?

To answer this empirically, I use a proxy for technology flows and regress it on immigration and other control variables. Technology flows are measured as the number of forward citations a US patent receives from inventors in a given sample country during a given year. The sample contains thirteen countries over the years 1995-2010. Given the characteristics of the data, a fixed-effects Poisson distribution model was applied to conduct the regression analysis.

The immigration was found to be positive and statistically significantly related to technology flows. The result is fairly robust for different regression specifications; all but one model show that the effect of immigration is statistically significant and all of the models show the effect to be positive. These results support the hypothesis that brain drain migration leads to technology flows back to the source country. Although my sample countries are considered economically developed, there is evidence to suggest they too suffer from brain drain migration to the US. Thus, the results found are significant and relevant for the sample countries analyzed in the paper.

Table of Contents

List of Tables

Chapter 1. Introduction

Immigration is a major aspect of globalization. As the world becomes increasingly integrated, it becomes important to learn more about the effects of this process. More specifically, what is the correlation between immigration and economic growth? And if immigration does create growth, what are the long-run implications for both the source country and the host country? The majority of academic literature in economics aimed at understanding the relationship between immigration and growth has focused mainly on immigration's effects on labor markets. The most basic labor market model shows that while immigration creates both gains and losses for different groups of people in the source and host countries, the net gain in real income as measured by GDP is positive. Moreover, the gain in output from the host country is greater than the loss of output in the source country. So immigration increases *total* output.

Though immigration has been extensively proven to increase net output, there is much more to the story, so to speak. That is, there are many sources of increased output, or economic growth. Very generally, there are three main causes of economic growth. The first is an increase in resources. The second is an increase in the quality of resources. Finally, the third is technological advance. According to Robert Solow's long run growth model, technological advance is the only form of growth sustainable in the long run.

In Solow's growth model, economies converge to a steady-state growth rate in the long run. The growth rate is achieved once the economy reaches its steady state level of capital. Various endogenous factors, including an increase in the labor force via immigration, can change the steady state level of capital and the rate at which an economy arrives there. Once there, however, the economy returns to its steady-state growth rate. If increases in the long run, steady-state growth rate are to be achieved, technological advance must occur. Thus, the correlation between immigration and technological advance is vital if we are interested in learning the effects of immigration on the long run, steady-state growth rate. That is, we must determine how sustainable this growth created by immigration is.

In labor economics, the economic growth created by immigration comes from a more efficient distribution of resources. Labor moves toward the country where its marginal product is greater. This growth, however, is caused by a one-time increase in the supply of labor in the host country. Thus, the basic model suggests this growth is created by the first cause: an increase in resources. That is, the labor movement creates a one-time redistribution of resources that is more efficient. As Solow indicates, however, this type of growth is not sustainable in the long run. Labor economics does not address any possible technological advance created by immigration. Therefore, we need to look outside the realm of traditional labor market models of immigration to see if immigration can create *sustainable* economic growth.

The effects of immigration reach far beyond just the labor market. Immigrants bring with them much more than labor supply. Just as each individual has a stock of knowledge, a unique set of preferences, and specific cultural characteristics, so too do immigrants. An immigrant population can change the composition of the entire economy, not just the labor market. Their stock of knowledge can greatly influence the ability of a society to innovate or advance technologically. Their preferences can greatly affect aggregate demand and the composition of goods produced in a society. Finally, their cultural characteristics can affect everything from the form of government to the religious makeup of a society. Thus, as mentioned before, labor market models cannot wholly capture the total effects of immigration.

This paper looks beyond the labor market to investigate the relationship between migration and the flow of knowledge or technology. The investigation, however, is a complicated process. Knowledge flows are notorious for being immeasurable; they leave no paper trail with which to capture a flow. Recently, though, patent citations have been used to capture technology flows. Just as references in an academic article cite previous knowledge upon which that article builds, so to do patents. A patent that cites a previous patent represents a flow of knowledge from the inventor of the cited patent to the inventor of the subsequent patent. Though these forward citations exist, they are incredibly difficult to find on an international scale.

The Patent Cooperation Treaty (PCT), concluded in 1970, made great strides in providing transparency of patents' information. For example, any patent application from a member state of the PCT is required to include citations of all previous works of art, including patents, upon which this new patent builds. Though patent protection can only be offered within a country, the PCT requires the search for all previous works of art to be performed on an international scale. Thus, an inventor seeking to patent in a specific country, for example, must cite all previous works of art, including those from other countries.

Even with the documentation of forward citations on an international scale, there are many difficulties associated with using patent citations to measure international technology flows. The patent application process is lengthy, so the timing of technology flows becomes ambiguous. To make matters worse, many countries are experiencing incredible delays between the time a patent application is filed and the time a patent is granted. This delay could substantially influence results of any empirical analysis. In addition, the documentation of international forward citations is not widely available. Many patent search engines, maintained and operated by various patent offices, contain only forward citations of national patents. Even those that contain international information have limited countries included in their database. Thus, data restrictions quickly become cumbersome for any empirical analysis. Because of the reasons listed above, virtually no literature exists investigating the correlation between international migration and international technology flows.

This paper looks at migration into the United States and technology flows from the US to other countries. The sample of countries included in my empirical analysis includes countries from Western Europe plus Japan and Australia. Because my sample consists of mainly economically advanced countries, information on forward citations contained in patents from inventors in these countries was available. Additional research revealed that these countries were experiencing brain drain migration to the US. Described in detail later in the paper, brain drain migration is the process whereby highly skilled labor migrates to another country in search of better economic opportunities. This immigration has been proven to harm the economies of the source countries and to benefit host countries. However, if this migration leads to technology flows from the host country back to the source countries, as my hypothesis later states, this process may create technological advance in source countries.

The paper begins by explicitly stating the hypothesis. Next, a review is made of the existing literature on the topic of the correlation between migration and technology flows. Then, the methodology, the econometric model, and the data are described. Following the data section, the initial results of the regressions are presented and discussed. After reviewing the initial results, a new model is formulated using revised definitions of the original variables as well as additional variables. The results of the regressions using the reformulated model are then discussed. Next, two different sensitivity analyses are performed on the revised model. The first analysis involves varying the immigration variable in the model; the second involves varying the regressors in the model. I then consider the results of a new panel data model disregarding the individual patent effect that was assumed in the previous models, but attending to the possible country and time effects. Finally, all results are discussed and future additions to the hypothesis, the model and the paper are explored.

1.1 Hypothesis

The hypothesis of this paper is that migration leads to knowledge flows from host countries to source countries. In this paper, the hypothesis is tested empirically on an international scale, using patent citations as a proxy for technology flows. Specifically, my hypothesis is that knowledge flows *from* the US are correlated with migration flows *to* the US. That is, migration to the US from other countries creates channels whereby information, or knowledge, flows back to immigrants' countries of origin or prior residence.

Chapter 2. The Literature

Before proceeding to the empirical estimation of the effects of immigration on technology flows, we must first survey the existing literature on this relationship. While the literature regarding immigration and its effects on economic growth are extensive, the literature on knowledge flows is quite limited. We must first discuss the literature focusing on a process known as the brain drain and its effects on the economies involved. Next we review the papers studying the effects of brain drain on the global agglomeration process. We then look at the studies conducted on the effects of both brain drain and agglomeration on source countries. Specifically, we review those studies that suggest some benefits to source countries may exist in the form of knowledge flows from host countries to source countries. Finally, we look at the few papers using patent citation data to empirically measure knowledge flows.

2.1 Brain Drain

We begin by examining the somewhat robust literature regarding the relationship between immigration and increases in the quality of the labor force. In other words, we examine whether immigration creates overall improvements in human capital. One common topic within this literature is that of the determinants and effects of brain drain. Brain drain is the process whereby highly skilled workers migrate from developing to developed countries. While the determinants of brain drain are vast, these migration flows are often motivated by greater earnings possibilities in the host country in the form of greater demand for skilled labor and thus higher paid employment opportunities.

In a paper titled "'Human Capital Flight': Impact of Migration on Income and Growth", authors Nadeem U. Haque and Se-Jik Kim use an endogenous growth model to examine the effects of brain drain on the host country and the source country. They find that brain drain will lead to a permanent increase in income and growth in the host country relative to the source country. Though the neoclassical approach predicts that human capital flight can be welfare improving overall, externalities not accounted for in this approach could create substantial welfare losses in the source country, such as inefficiencies associated with a less diverse workforce. As a result of the brain drain, return on investment in human capital can actually be negative after a certain point in the source country. This means that the source country only has an incentive to invest in its native inhabitants' education up to a certain point or skill level, after which the inhabitants become more likely to emigrate, thus activating the brain drain process. The conclusion of this paper, like other papers regarding brain drain, is that the "brain gain" in the host country can be more than offset by the brain drain in the source country. Not only that, the brain drain process can create disincentives for countries to invest in education, another welfare-reducing effect of immigration.

Some literature, however, cites possible benefits to source countries in the presence of brain drain. In the paper "Scale Economies in Education and the 'Brain Drain' Problem", author Kaz Miyagiwa writes that physical distance still impedes the dissemination of technology and knowledge, and that spillovers are still restricted to relatively small geographic areas. Thus, the increasing returns to higher education made available through the agglomeration of skilled professionals create incentives for the highly skilled to "stay put". So in contrast to Haque and Kim, Miyagiwa argues that investing in higher education may not necessarily encourage brain drain. If brain drain occurs, however, Miyagiwa finds the same detrimental effects on the source country that Haque and Kim find. In addition, Miyagiwa claims that the aggregate income of the source country can decline in the face of brain drain, even when those that migrated to the host country are included. If this is the case, remittances of those who emigrated will not be adequate to sufficiently compensate those who stayed. Thus, Miyagiwa reaches the same conclusion that highly skilled workers should somehow be restricted or discouraged from emigrating.

2.2 Brain Drain and Agglomeration

Solow's neoclassical growth model indicates that technological advance is the only form of growth sustainable in the long run. Thus, it is imperative to review the existing literature devoted to investigating the correlation between immigration and technological advance or technology transfers. Much of this literature focuses on the respective levels of technology in the source country, the host country and the ensuing migration flows. In their paper "The Impact of Differences in Levels of Technology on International Labor Migration", Oded Galor and Oded Stark find that, other things equal, migration will flow from the technologically inferior country to the technologically superior country. This results from a higher return to the factors of production in the technologically superior country. Hitoshi Kondo finds the same result in the paper "International Factor Mobility and Production Technology". Thus, the empirical evidence suggests that immigration flows mainly in one direction: from developing to developed countries. This is a clear process of agglomeration. The next logical step would be to learn more about this process and how quickly it is occurring via migration.

Agglomeration is a process whereby objects collect into a single cluster or mass. In economics, it describes the tendency of factors of production to gather in specific geographic area or region. Somewhat paradoxically, it has been occurring against a backdrop of extensive globalization and increased global economic integration. Gianmarco Ottaviano and Diego Puga, authors of "Agglomeration in the Global Economy: A Survey in the 'New Economic Geography'", claim that this "area" of agglomeration can range from small industrial districts, such as the carpet production industry in Dalton, Georgia, to interstate and even international regions, such as the '"Manufacturing Belt" across the northern region of the US.

In the paper "Agglomeration and the Location of Innovative Activity", author David Audretsch explains why this agglomeration process occurs. According to Audretsch, innovative activity, or new knowledge, has recently become the leading source of comparative advantage among developed countries. One reason for this is the increased competition from emerging economies of the developing countries in Central Europe and Southeast Asia. Unlike traditional factors of production, knowledge does not spill over across large areas of geographic space. Moreover, physical proximity of different firms performing the respective steps of the production process is beneficial in that it increases efficiency and thus reduces costs.

Ottaviano and Puga add to the reasons behind agglomeration, citing that firms that locate near large markets can create economies of scale and minimize transactions costs. All of these characteristics create incentives to localize geographically. In other words, today's producers have an incentive to agglomerate into small geographic areas. These areas, of course, can provide better employment opportunities and higher wages to immigrants, thus creating the immigration flows toward technologically superior countries and the brain drain.

So how fast is the agglomeration process happening? How quickly are people, and other factors of production, moving to specific geographic areas? In his article, "The World Is Spiky", Richard Florida shows that migration from rural areas to cities has accelerated tremendously in the past two centuries. On average, 50% of the world's population currently resides in urban areas, up from 30% in 1950 and just 3% in 1800. This number jumps to as much as 75% of the population for advanced countries. In addition, Florida tries to capture the areas of the most innovation by measuring patents from resident inventors in over 100 nations. In 2002, 85% of patents recorded were given to residents of only 5 countries: Japan, the US, South Korea, Germany and Russia. This reveals very clearly that technological progress is indeed undergoing a process of agglomeration.

Juan Dolado, Alessandra Goria and Andrea Ichino have also written a paper on the evidence of agglomeration through immigration. In the paper, the immigration flows for 23 OECD countries are observed over the period 1960 – 1985. They find that population growth has become increasingly due to immigration over this period. For example: "If, on average, the population growth due to immigrants was 56% of the total population growth in the 60s, this percentage becomes 91% in the 70s and it climbs up to 111% in the 80s [meaning the population growth was greater than the growth in non-immigrant population]" (Dolado, Goria and Ichino, 1994). They also find that more immigration has led to greater human capital. That is, the immigrants, on average, were generally as skilled as or more skilled than the native population. Because the majority of OECD countries are developed countries, this again shows that migration toward advanced countries has grown in recent decades.

The findings in the aforementioned Galor paper and Stark and Kondo paper regarding migration toward technologically superior countries are simply extensions of the brain drain argument, and thus are not very surprising. In addition, these papers find that differences in technology cause migration and the agglomeration process. The more interesting relationship to investigate, however, would be causation in the other direction. That is, does migration help close gaps in technology between countries? Can source countries "catch-up" to host countries via dissemination of knowledge from the host country?

2.3 Technology Flows and Source Countries

According to the literature, there are several avenues through which migration can indeed send knowledge from developed countries back to source countries. AnnaLee Saxenian, author of "Brain Circulation: How High-Skill Immigration Makes Everyone Better Off", argues that we should really start looking at brain drain as "brain circulation" because high-skilled immigration can benefit the source country in addition to benefitting the host country. In her article, Saxenian uses the case of Silicon Valley to show how immigrants in developed countries can support their counterparts at home. According to Saxenian, the numerous ethnic groups, who account for an increasing number of the Valley's highly skilled workers, have formed social and professional networks with one another to share information and expedite innovation. The transnational networks have, in essence, created a platform for globalizing their technology firms that started in Silicon Valley. Members of these networks are able to serve as middlemen that link businesses in Asia and other distant areas with those in the US. For example, Silicon Valley's Asian engineers have built strong connections with technology communities in India and Taiwan. The experience of Silicon Valley reveals that highly skilled immigrants are now maintaining relationships with their professional colleagues at home, creating information flows back to the source country.

The NBER working paper "Gone but Not Forgotten: Labor Flows, Knowledge Spillovers, and Enduring Social Capital", by Ajay Agrawal, Iain Cockburn and John McHale, provides more evidence of bidirectional technology flows. Like "The World is

Spiky" article, this paper uses patents to look at technological advance and innovation. The paper finds that knowledge flows more strongly to prior locations of inventors. This reveals that social and professional ties between highly skilled immigrants and their associates from their native countries facilitate some form of knowledge transfer even after the individuals are separated via migration of the former. The paper finds these spillovers particularly strong in technology fields, where transferring knowledge can be more costly.

Another way in which immigrants can send technology back to the source countries is through return migration. If the previous two papers are correct in showing that technology does, in fact, flow back to source countries, then the source countries will begin to grow. This is currently the case in Southeast Asia. As these source countries develop, new lucrative employment opportunities for the high-skilled labor that previously emigrated will emerge, drawing these immigrants homeward. According to the NBER working paper "Return Migration as a Channel of 'Brain Gain'", by Karin Mayr and Giovanni Peri, the return migration channel is a significant factor in reversing the welfare-reducing effects of brain drain and turning them into a "brain gain" for the source country. In addition, this paper provides empirical evidence that highly skilled immigrants are increasingly migrating temporarily, bringing back with them, of course, the knowledge they acquired from abroad. From these three papers, it is clear that technology can flow back to the sending country, revealing that both host and source countries can benefit from immigration.

2.4 Technology Flows and Patent Data

While the above literature suggests that some work is indeed being done in examining the relationship between migration and technology flows, it is scarce and fairly one-dimensional. That is, little empirical work has been done on a large scale to study this relationship. The lack of extensive literature can be attributed to the difficulty with which technology flows can be measured. As Paul Krugman wrote in 1991, "…knowledge flows, by contrast, are invisible; they leave no paper trail by which they may be measured and tracked, and there is nothing to prevent the theorist from assuming anything about them that she likes".

Some knowledge flows, however, can be traced using patent citations. A patent is a monopoly over some piece of intellectual property for a certain period of time. It is granted to an inventor or applicant by a sovereign state, in most cases a country. Often times, a patent is an extension of previously patented technology. If so, that subsequent patent (or patent application) must cite the previous patent upon which it builds. The original patent will be denoted the *originating patent*, the subsequent patents that cite the originating patent will be denoted *citing patent*. Each patent document contains detailed information regarding the inventor, including their geographic location. If we can determine the location of the inventor of both the originating patent and the citing patent, we can obtain the path of knowledge flow – from the location of the inventor of the originating patent to the location of the inventor of the citing patent. Thus, patent citations can be used as a proxy for technology flows.

In their paper "Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations", Jaffe, Trajtenberg and Henderson use patent citation data to study the geographic location of R&D spillovers with the hypothesis that the spillovers are geographically localized. They begin with a group of originating patents within the US and Canada, then find all citing patents within the same region. A group of "control" patents are constructed using patents with the closest dates to those of the citing patents. The study finds that citing patents are more likely to come from the same geographic location as the cited patent, indicating that knowledge flows are indeed geographically localized.

The aforementioned NBER working paper "Gone but Not Forgotten: Labor Flows, Knowledge Spillovers, and Enduring Social Capital" uses the same methodology as Jaffe, Trajtenberg and Henderson to capture knowledge flows in the US and Canada. This paper, however, goes a step further to investigate the incidence of citing patents in prior locations of inventors. The paper finds that patents are cited disproportionately where the inventor receiving the patent previously resided, revealing that knowledge flows do result from migration.

While the previous two papers look at knowledge flows within the US and Canada, Sjöholm uses patent citations to study the flow of knowledge across borders in his paper "International Transfer of Knowledge: The Role of International Trade and Geographic Proximity". Sjöholm measures trade flows from Sweden and inspects Swedish patents to find knowledge flows from other countries into Sweden. Employing a conditional logit model, Sjöholm finds that trade has a positive and statistically significant effect on knowledge flows.

2.5 Conclusion of Literature

After reviewing the literature on immigration and sustainable economic growth, it is clear that migration patterns, their determinants and their effects are not simple; they are not linear and they are not static. When looking specifically at the correlation between immigration and technological progress, it very quickly becomes clear that the relationship is complex and bidirectional. Technological progress has a distinct and real effect on immigration, but immigration flows can also affect technological progress, or at least the dissemination of knowledge. The majority of the literature seems to focus more on technologically superior countries attracting immigration, which in essence is merely the process of agglomeration and the brain drain. Because there are both gains and losses associated with brain drain, however, it is important to ask: who wins and who loses? And can the winners sufficiently compensate the losers? If technology flows back to a source country, they need not suffer from the welfare-reducing effects of brain drain. In fact, they could benefit from sending labor abroad if it meant expedited technology transfers from developed countries. As mentioned before, however, this process remains relatively untouched in the field of economics. That is, little is known about the effects of immigration on technological progress in the source country.

Chapter 3. The Methodology

3.1 Patents and Patent Citations

In order to investigate the correlation between technology flows and migration, one first needs a method in which to measure the technology flow. As a proxy for this, one can use patent citations. A patent is a monopoly over some piece of intellectual property for a certain period of time. It is granted to an inventor or applicant by a sovereign state, in most cases a country. The World Intellectual Property Organization (WIPO) defines a patent as "…an exclusive right granted for an invention, which is a product or a process that provides, in general, a new way of doing something, or offers a new technical solution to a problem." In order for an invention to be profitable, it must fulfill various conditions. The invention must be novel, meaning that it contains some new characteristic which is not already known in the body of existing knowledge, known as "prior art". Once a patent is granted, a document is created that contains information about the inventor, the inventor's employer, and an extensive description of the invention. This patent document is considered public information, and is organized by a classification system in order to be searchable. The reason for this intricate classification system is that, often times, a patent is an extension of previously patented technology. At some point during the patent process, a patent examiner must perform a search in order to find any prior art upon which the patent builds. If prior art is found, it must be cited. This citation represents a flow of technology or knowledge from the inventor of the prior art to the inventor of the current patent upon which the search is performed.

Currently, patents can only be granted by and protected in countries, but not internationally. However, the patent examiner is responsible for consulting databases that contain information on patents worldwide. Thus, the "prior art" being searched is not confined just to patents granted in the country where the inventor is applying for a patent. This is relevant to my research because I aim to capture international flows of knowledge, not flows within countries.

While patents are currently only granted in specific countries, there are steps being taken to streamline the application process so that an inventor may apply for a patent in more than one country simultaneously. The Patent Cooperation Treaty (PCT) was signed in 1970 at the Washington Diplomatic Conference on the Patent Cooperation Treaty, and has been modified since in 1979, 1984 and 2001. The PCT is an international patent law treaty aimed at providing a unified procedure and legal structure for the patent application process across countries. The first of these treaties was the Paris Convention for the Protection of Industrial Property, signed in Paris in 1883. This treaty was the first to establish a union for the protection of intellectual property. Any contracting member of this union is eligible to become a member of the PCT. As of 2011, there were 174 contracting member countries to the Paris Convention for the Protection of Industrial Property, and there are currently 145 contracting member countries to the PCT with the country of Brunei Darussalam becoming the 145th contracting member on April 24th, 2012.

Under the PCT, a national or resident of any of the 145 contracting states may seek patent protection for an invention in each of the contracting states concurrently by filing an international patent application. Though this application itself will not enable patent protection in each of the contracting states, it allows the inventor to take several steps in the application all at once, as opposed to undergoing the process in each state individually. For example, all international patent applications become subject to an international search to be carried out by a member of the International Searching Authority (ISA). Like the aforementioned searches, the purpose of the international search is to find prior art upon which the invention builds. The ISA then publishes their findings, including all citations of relevant prior art, in a document called an international search report. This report is taken into consideration by national patent authorities when the patent applicant enters the national phase of the application process, when the patent is sought in specific countries. Some national patent authorities will rely solely on this report, deeming it unnecessary to perform supplementary searches and saving the applicant time and fees to be paid for searching and translation.

In addition to the PCT, there are numerous regional offices that will assist in applying for patent protection throughout the whole region. The European Patent Office (EPO) is one such regional patent office. Created October 7, 1977, the EPO is responsible for granting European patents and conducting search reports for patent applications submitted to various national patent offices across Europe. The EPO consists of 38 member states throughout Europe. The patents the EPO grants are not "international" patents, but rather a bundle of national patents. Another prominent regional patent office is the African Regional Intellectual Property Organization (ARIPO). The foundation for ARIPO was laid in 1976 when an agreement on the creation of the Industrial Property Organization for English-Speaking Africa (ESARIPO) was signed in Lusaka, Zambia.

The purpose of the organization was to pool resources of member countries together concerning intellectual property matters. These regional and international patent offices not only streamline the patent application process for inventors, but they also make available information on patent citations across countries, information that is integral to my research.

3.2 Forward Patent Citations

From the previous section we see that patents are necessarily cited whenever subsequent inventions build upon them. While the citation of the original or originating patent appears in the later patent document as a "cited document", the citation of the subsequent patent may also appear on the originating patent as a "citing document". These citations of later patents are called forward citations, and are searchable via some databases. As previously mentioned, these forward citations represent a flow of knowledge or technology from the inventor of the originating patent to the inventor of the forward citation. In this paper, I am interested in obtaining flows from the US to other countries. That is, I am interested in finding all patents from inventors in foreign countries that have cited US patents granted to inventors residing in the US. It is important to note that I am looking for US patents from inventors from the US, not simply US patents. This is because a large portion of US patents are granted to foreigners. According to Jaffe and his colleagues, this portion was approximately 40 percent. Thus, I begin with a sample of US patents and find all the forward citations from foreign inventors.

Patenting activity in the US is immense. In 1998 alone, 163,204 patents were granted. Because it is necessary to look up each patent individually to find its forward citations, I must choose a significantly reduced sample of US patents. Patents in the US are classified using the US Patent Classification System, maintained by the US Office of Patent Classification. There are currently approximately 987 "parent" US Patent Classes and 35 Patent Classifications for "design patents". I have chosen to use a sample of US patents from US Patent Class 47: Plant Husbandry. Plant Husbandry is defined by the US Patent and Trademark Office (USPTO) as " … the parent class for apparatus and processes employed in treating the earth and its products and includes all inventions relating thereto that have not been especially provided for in other classes." This patent class contains 89 subclasses, which were all included in the sample. I use this particular classification because it contains the most agricultural patents. Agricultural products account for a large portion of trade for the countries included in my sample. In addition, advances in agricultural would be highly beneficial to developing countries, as many rely on agricultural as a main source of income and sustenance. I find all patents in this class granted to inventors from 1998 to 2002. I use this date range because Jaffe and his colleagues have suggested that the average citation lag – the time it takes for a patent to be forward cited, was somewhere between two and six years. I wanted to avoid disturbances in patent activity due to the financial collapse of 2008, so I use a sample that ends in 2002 – allowing at least six years of relative international economic prosperity in which to apply for and cite previous patents.

To construct this sample, I use a database run by the USPTO called the Patent Full-Text and Image Database. I perform an advanced search using the following criteria:

"ISD/1/1/1998->12/31/2002"

ISD –Issue Date – "This field contains the date the patent was officially issued by the US Patent and Trademark Office." The data range searched was from January 1st, 1998 to December 31st, 2002.

"CCL/47/$"

CCL – Current US Classification – "This field contains the original and cross-reference US Classification(s) to which the published application was assigned at the time of publication. This field includes both primary and secondary class information." The classification searched was US Class 47, Plant Husbandry, and all subclasses.

"IS/AK"

IS – Inventor State – "This field contains the US state of residence of the inventor at the time of publication." Because I needed only patents granted to Inventors residing in the US, I searched and compiled patents with an inventor state of each of the US states.

After searching all patents for the US classification of 47, including all subclasses, there were a total of 1366 US patents. Some of these patents, however, were duplicates, as patents could have multiple inventors from multiple states. I removed the duplicates after finding all forward citations of these patents. I discuss the findings below.

Each of these US patents were sought individually on the European Patent Office's (EPO) database, Espacenet[1], using their US patent numbers as search guides. Each US patent document on Espacenet contains information on "citing documents", which includes any of the aforementioned forward citations. Information was documented on all of the citing documents filed by inventors from countries other than the US. The following information was documented for each forward citation fitting the aforementioned criteria:

[1] For a full description of Espacenet and the international application organizations, see Appendix A.

Country Code – “Country codes are two letters indicating the country or organisation where the patent application was filed or granted (eg GB).”

Inventor Country Code – The country code next to the inventor listed on the patent/patent application. This stands for the country of residence of the inventor, not of citizenship. This information is provided by the inventor or applicant filling out the application.

Applicant – “An applicant is a person or organisation (e.g. company, university, etc.) who/which has filed a patent application. There may be more than one applicant per application.”

Applicant Country Code – The country code next to the applicant listed on the patent/patent application. This stands for the country of residence of the applicant, not of citizenship. This information is provided by the inventor of applicant filling out the application.

Publication Number – “The publication number is the number assigned to a patent application on publication. Publication numbers are generally made up of a country code (two letters) and a serial number (variable, one to twelve digits) (eg DE202004009768).”

Publication Date – “The publication date is the date on which the patent application was first published. It is the date on which the patent document is made available to the public, thereby becoming part of the state of the art.”

Priority Number – “The priority number is the number of the application in respect of which priority is claimed, i.e. it is the same as the application number of the claimed priority document.”

For the US patents I was searching, I also documented the following information: applicant, applicant country code, publication date, and priority number. For both originating and forward citing patents, I noted whether the applicant and the inventor were the same. Though I have not yet used this information in my empirical research, I believe it may hold interesting insight into the dynamic of technology flows.

After searching all US patents and documenting all relevant information, it was necessary to remove a substantial amount of forward citations due to the lack of resources included in the Espacenet database. That is, only information from a certain number of countries' own patent offices are contained in Espacenet (for simplicity's sake, these countries will be called member countries). For example, the US is a member country. This means that information from the USPTO is included in the search engine. Thus, Espacenet will have documentation of forward citations for US patents, including inventors worldwide who have sought patent protection in the US, any of the other member countries, or in any of the international patent application organizations included in the Espacenet database[2]. However, China, for example, is not a country whose patent office's information is included in Espacenet. Thus, no forward citations included in Chinese patents will be revealed through an Espacenet search. In other words, forward citations from Chinese inventors will only be found on Espacenet if these inventors are applying for patents in one of the member countries or international patent application organizations. It is fairly easy to assume that a large amount of Chinese patents will come from Chinese inventors. Thus, a large amount of forward citations from Chinese

---

[2] For a list of "member countries" and international patent application organizations included in the dataset, see Appendix B.

inventors will not be revealed via an Espacenet patent search. So, it would not be wholly representative to include only forward citations from Chinese inventors seeking patent protection in member countries or the aforementioned international patent application organizations. Therefore, I included forward citations from inventors from only sample countries during the time period in which I sampled US patents.

In the end, I include in my sample only forward citations from inventors who reside in one of the sample countries and patent in one of the member countries or the international patent organizations included in Espacenet. It can be easily seen that inventors from the countries included in my sample will file patents in their home countries, neighboring countries, one of the various international patent application offices, or the US – as all of these are included in the Espacenet database, I argue that my sample has captured the vast majority of forward citations from inventors in the sample countries.

In addition, I only include one forward citation for each inventor per US patent. This is because it is possible for one inventor to use the information from a US patent to create several new inventions, and I aim to measure the initial transfer of technology, not the number of times the inventor uses this information. Meaning, I do not aim to measure how many times this technology is used after the transfer is made. The reason for this is the restriction of the data; if a patent is cited by an inventor or an applicant more than a certain number of times, this information is not shown on the results page of an Espacenet patent search. More detailed information on the composition of forward citations in the sample, including a list of sample countries, is provided in the Data section.

Chapter 4. The Model

The next step taken was to develop an appropriate econometric model with which to effectively evaluate the possible correlation between migration and technology flows. I use forward citations, the proxy for technology flows, as the dependent variable, and look at independent variables that could affect the probability that these citations occur. That is, the independent variables explain how many times a US patent is cited by an inventor from a sample country in a given year.

4.1 The Fixed Effects Regression Model

Given the characteristics of the data, the most appropriate econometric model for my analysis is the fixed effects (FE) Poisson regression model. More specifically, the dependent variable in my regression is a nonnegative count variable with no theoretical upper bound; it takes on integer values greater than or equal to zero. Thus, a parametric model that predicts expected values only greater than zero is necessary. Because a FE pooled OLS model is a linear model, it is possible to get values of $\mathbf{x}$ where $\mathbf{x}\hat{\boldsymbol{\beta}}_{FE} < 0$. Thus, a FE pooled OLS model is inappropriate to use for my dataset.

However, I begin by estimating a FE pooled OLS model, simply to examine the regression coefficients as a reference point for future regressions. Fixed effects models are used to control for heterogeneity introduced by some unobservable, time-invariant individual effect that is correlated with the regressors in the model. In this model, the unobservable fixed effect captures the unobservable individual characteristics of each of the US patents that may affect the amount of forward citations each patent receives, other things equal. For example, “high tech” patents tend to be cited much more often than other patents. Likewise, patents representing a higher quality of knowledge presumably would be cited more often that other patents. In addition, I argue this effect if correlated with the regressors in the model, specifically immigration. Immigrants from a certain country or a certain time period may be more skilled or skilled in “higher tech” industries, and thus may be more likely to produce “high tech” patents that would be cited more often. This would cause the unobservable effect to be correlated with immigration. Below, the FE pooled OLS estimator is derived using a general model; and a more specific model is described later[3].

Consider the following linear model for T time periods:

$$y_{it} = \mathbf{x}_{it}\boldsymbol{\beta} + c_i + u_{it},\ \ t = 1, 2, \dots, T; i = 1, 2, \dots, N \quad (1)$$

where c_i is an unobserved, time invariant effect associated with each US patent. In addition, $\mathbf{x}_{it}$is the vector of independent variables associated with patent i at time t. Averaging the data over time periods $t = 1, 2, \dots, T$, gives the cross section equation:

$$\bar{y}_i = \bar{\mathbf{x}}_i\boldsymbol{\beta} + c_i + \bar{u}_i,\ \ i = 1, 2, \dots, N \quad (2)$$

where $\bar{y}_i = \sum_{t=1}^{T} {y_{it}}/{T}$, $\bar{\mathbf{x}}_i = \sum_{t=1}^{T} {\mathbf{x}_{it}}/{T}$, and $\bar{u}_i = \sum_{t=1}^{T} {u_{it}}/{T}$. Subtracting equation (1) from equation (2), gives the FE transformed cross section equation:

$$y_{it} - \bar{y}_i = (\mathbf{x}_{it} - \bar{\mathbf{x}}_i)\boldsymbol{\beta} + u_{it} - \bar{u}_i,\ \ t = 1, 2, \dots, T; i = 1, 2, \dots, N$$

or

$$\ddot{y}_{it} = \ddot{\mathbf{x}}_{it}\boldsymbol{\beta} + \ddot{u}_{it},\ \ t = 1, 2, \dots, T; i = 1, 2, \dots, N \quad (3)$$

where $\ddot{y}_{it} \equiv y_{it} - \bar{y}_i$, $\ddot{\mathbf{x}}_{it} \equiv \mathbf{x}_{it} - \bar{\mathbf{x}}_i$, and $\ddot{u}_{it} \equiv u_{it} - \bar{u}_i$. Notice that this procedure has removed the unobserved effect c_i.

The FE estimator, $\widehat{\boldsymbol{\beta}}_{FE}$ (referred to as the within estimator), is obtained by using pooled OLS to estimate equation (3):

---

[3] This derivation is based on that of StataCorp (2009) and Wooldridge (2002).

$$\widehat{\boldsymbol{\beta}}_{FE} = \left(\sum_{i=1}^{N} \ddot{\mathbf{X}}_i{}'\ddot{\mathbf{X}}_{\boldsymbol{i}}\right)^{-1} \left(\sum_{i=1}^{N} \ddot{\mathbf{X}}_i{}'\ddot{\mathbf{y}}_{\boldsymbol{i}}\right) = \left(\sum_{i=1}^{N}\sum_{t=1}^{T} \ddot{\mathbf{x}}_{it}{}'\ddot{\mathbf{x}}_{\boldsymbol{it}}\right)^{-1} \left(\sum_{i=1}^{N}\sum_{t=1}^{T} \ddot{\mathbf{x}}_{it}{}'\ddot{y}_{\boldsymbol{it}}\right)$$

Under only the assumption of strict exogeneity of the explanatory variables conditional on the fixed effect, c_i: $\mathrm{E}(y_{it}|\mathbf{x}_{i1}, \ldots, \mathbf{x}_{iT}, ci) = \mathrm{E}(y_{it}|\mathbf{x}_{it}, c_{it})$, and the standard rank condition on the explanatory variables from equation (3):rank$\left(\ddot{\mathbf{X}}'\ddot{\mathbf{X}}\right) = K$, $\widehat{\boldsymbol{\beta}}_{FE}$ is an unbiased estimator conditional on **X.** The model thus allows $E(c_i|\mathbf{x}_i)$ to be any function of $\mathbf{x}_i$. In other words, the model allows arbitrary correlation between the unobservable c_i and $\mathbf{x}_{it}$. However, this requires the exclusion of any time invariant factors in $\mathbf{x}_{it}$ (unless they are interacted with the time variant factors), as the effects of these would be indistinguishable from c_i.

4.2 The Fixed Effects Poisson Regression Model

The dependent variable, y, in this model is the amount of forward citations that occur from an inventor in a given sample country in a given year. As mentioned in the previous section, it is a nonnegative count variable with no theoretical upper bound; it takes on integer values greater than or equal to zero. Thus, the correct method of estimation will produce predicted values of y that are nonnegative. If the FE model described above model is estimated using pooled OLS with $\hat{\boldsymbol{\beta}}_{FE}$ being the FE estimator, it is possible to get values of $\mathbf{x}$ where $\mathbf{x}\hat{\boldsymbol{\beta}}_{FE} < 0$. Thus, pooled OLS is not the appropriate estimation method, and $\hat{\boldsymbol{\beta}}_{FE}$ is not an efficient estimator. Log-linearizing the data and continuing with OLS is often appropriate for strictly positive variables, but only if the dependent variable is non-zero. Because my dependent value takes on the value of zero for a non-trivial portion of the dataset, this approach is not possible. Another approach entails using nonlinear least squares (NLS) to estimate the model. However, NLS is only efficient under the condition of homoskedasticity. Because the distributions of count data often imply heteroskedasticity, this method is not ideal.

The most popular model for count data is the Poisson regression model; if the independent variable given under $\mathbf{x}$ is distributed as Poisson, the conditional maximum likelihood estimators derived from the Poisson density function are fully efficient. The fixed effects (FE) Poisson regression model, developed by Hausman, Hall, and Griliches (1984), is the most appropriate model to estimate. This model is estimated, rather obviously, using FE Poisson estimation, which is a conditional maximum likelihood estimation (MLE) technique. The derivation of the conditional log likelihood function and $\widehat{\boldsymbol{\beta}}_{FEP}$ follows. Again, a more general model is used in the derivation; a more specific model is described later[4].

Consider the following density function for T time periods. Let the conditional mean, $\mathrm{E}(y_t|\mathbf{x}_t, c) = cm(\mathbf{x}_t, \boldsymbol{\beta})$ where $c = \exp(a)$ and $m(\mathbf{x}_t, \boldsymbol{\beta}) = \exp(\mathbf{x}_t, \boldsymbol{\beta})$. If y given under $\mathbf{x}$ is distributed as Poisson:

$$f(y_{it}|\mathbf{x}_{it}, c_i) = \mathrm{P}(\mathrm{Y}_{it} = y_{it}|\mathbf{x}_{it}, c_i) = \exp\{-exp(a_i + \mathbf{x}_{it}\boldsymbol{\beta})\} \exp(a_i + \mathbf{x}_{it}\boldsymbol{\beta})^{y_{it}} / y_{it}!$$

$$= \frac{1}{y_{it}!} \exp\{-\exp(a_i) \exp(\mathbf{x}_{it}\boldsymbol{\beta}) + a_i y_{it}\} \exp(\mathbf{x}_{it}\boldsymbol{\beta})^{y_{it}}$$

where $t = 1, 2, \ldots, T; i = 1, 2, \ldots, N$.

Because we assume exogeneity: $\mathrm{E}(y_t|\mathbf{x}_1, \ldots, \mathbf{x}_T, c) = \mathrm{E}(y_t|\mathbf{x}_t, c)$, the joint probability density function within a panel can be written as:

$$f(y_i|\mathbf{X}_i, c_i) = \mathrm{P}(\mathrm{Y}_{i1} = y_{i1}, \ldots, \mathrm{Y}_{iT} = y_{iT}|\mathbf{X}_i, c_i)$$

$$= \prod_{t=1}^{T} \frac{1}{y_{it}!} \exp\{-\exp(a_i) \exp(\mathbf{x}_{it}\boldsymbol{\beta}) + a_i y_{it}\} \exp(\mathbf{x}_{it}\boldsymbol{\beta})^{y_{it}}$$

$$= \left(\prod_{t=1}^{T} \frac{\exp(\mathbf{x}_{it}\boldsymbol{\beta})^{y_{it}}}{y_{it}!}\right) \exp\left\{-\exp(a_i) \sum_t \exp(\mathbf{x}_{it}\boldsymbol{\beta}) + a_i \sum_t y_{it}\right\}$$

where $i = 1, 2, \ldots, N$.

---

[4] This derivation is based on that of Hausman, Hall, and Griliches (1984), StataCorp (2009) and Wooldridge (2002).

The sum across time of the Poisson independent random variables within a panel is distributed as a Poisson, each with the conditional mean:

$\sum_t \mathrm{E}(y_{it}|\mathbf{x}_{it}, c_i) = \sum_t c_i m(\mathbf{x}_{it}, \boldsymbol{\beta})$. So:

$$\mathrm{P}\left(\sum_t \mathrm{Y}_{it} = \sum_t y_{it} \,|\mathbf{X}_{i,} c_i\right) =$$

$$\frac{1}{(\sum_t y_{it})!}\exp\left\{-\exp(a_i)\sum_t \exp(\mathbf{x}_{it}\boldsymbol{\beta}) + a_i \sum_t y_{it}\right\}\left\{\sum_t \exp(\mathbf{x}_{it}\boldsymbol{\beta})\right\}^{\sum_t y_{it}}$$

So the conditional likelihood function is obtained using a joint probability distribution conditional on the sum of outcomes across t:

$$\mathrm{P}\left(\mathrm{Y}_{i1} = y_{i1}, \ldots, \mathrm{Y}_{it} = y_{it}|\mathbf{X}_i, c_i, \left(\sum_t \mathrm{Y}_{it} = \sum_t y_{it} \,|\mathbf{X}_{i,} c_i\right)\right) =$$

$$\frac{\left[\left(\prod_{t=1}^{T} \frac{\exp\,(\mathbf{x}_{it}\boldsymbol{\beta})^{y_{it}}}{y_{it}!}\right) \exp\{-\exp(a) \sum_t \exp(\mathbf{x}_{it}\boldsymbol{\beta}) + a_i \sum_t y_{it}\}\right]}{\frac{1}{(\sum_t y_{it})!}\exp\{-\exp(a_i)\sum_t \exp(\mathbf{x}_{it}\boldsymbol{\beta}) + a_i \sum_t y_{it}\}\{\sum_t \exp(\mathbf{x}_{it}\boldsymbol{\beta})\}^{\sum_t y_{it}}}$$

$$= \left(\sum_t y_{it}\right)! \prod_{t=1}^{T} \frac{\exp\,(\mathbf{x}_{it}\boldsymbol{\beta})^{y_{it}}}{y_{it}!\,\{\sum_{r=1}^{T} \mathbf{x}_{ir}\boldsymbol{\beta}\}^{y_{it}}}$$

Notice the above equation does not depend on $c = \exp(a)$. The conditional log likelihood is thus given by:

$$l = \log \prod_{i=1}^{N} \left[\left(\sum_t y_{it}\right)! \prod_{t=1}^{T} \frac{\exp\,(\mathbf{x}_{it}\boldsymbol{\beta})^{y_{it}}}{y_{it}!\,\{\sum_r \exp(\mathbf{x}_{ir}\boldsymbol{\beta})\}^{y_{it}}}\right]$$

$$= \log \prod_{i=1}^{N} \left\{ \frac{(\sum_t y_{it})!}{\prod_{t=1}^{T} y_{it}!} \prod_{t=1}^{T} p_{it}^{y_{it}} \right\}$$

$$= \sum_{i=1}^{N} \left\{ \log \Gamma\left(\sum_{t=1}^{T} y_{it} + 1 \right) - \sum_{t=1}^{T} \log \Gamma\,(y_{it} + 1) + \sum_{t=1}^{T} y_{it} \log p_{it} \right\}$$

where $p_{it} = {\exp^{\mathbf{x}_{it}\boldsymbol{\beta}}} / {\sum_r \exp^{\mathbf{x}_{ir}\boldsymbol{\beta}}}$

$\widehat{\boldsymbol{\beta}}_{FEP}$ will be defined as the estimator that maximizes the terms in the conditional log likelihood function that depend on $\boldsymbol{\beta}$:

$$l(\boldsymbol{\beta}) = \sum_{i=1}^{N} \sum_{t=1}^{T} y_{it}[\log p_{it}]$$

That is, $\widehat{\boldsymbol{\beta}}_{FEP}$ will be chosen to solve the following equation:

$$\textstyle\sum_{i=1}^{N} \left({\partial l_i(\widehat{\boldsymbol{\beta}}_{FEP})} / {\partial \widehat{\boldsymbol{\beta}}_{FEP}} \right) = 0$$

where

$${\partial l_i(\widehat{\boldsymbol{\beta}}_{FEP})} / {\partial \widehat{\boldsymbol{\beta}}_{FEP}} = \sum_{t=1}^{T} y_{it} \left[{\left(\frac{\partial p_{it}}{\partial \widehat{\boldsymbol{\beta}}_{FEP}} \right)'} / {p_{it}} \right]$$

This estimation method has the attractive robustness property that, under only the aforementioned assumption of exogeneity, the fixed effects Poisson (FEP) estimator, $\widehat{\boldsymbol{\beta}}_{FEP}$ is consistent. As with $\widehat{\boldsymbol{\beta}}_{FE}$, the FE estimator obtained using pooled OLS, the model cannot contain any time invariant factors in $\mathbf{x}_{it}$. The model does allow for overdispersion or underdispersion, which occur when $\sigma^2 \neq 1$ in the following equation relating the conditional variance to the conditional mean, referred to as the Poisson generalized linear models (GLM) assumption: $\mathrm{Var}(y|\mathbf{x}) = \sigma^2\mathrm{E}(y|\mathbf{x})$ [this is a weaker version of the Poisson variance assumption: $\mathrm{Var}(y|\mathbf{x}) = \mathrm{E}(y|\mathbf{x})$]. Overdispersion occurs when $\sigma^2 > 1$, meaning the variance is greater than the mean. This would result in a report of standard errors that are too small, and any hypothesis testing conducted using these would be inaccurate. Underdispersion, which is less common than its counterpart, occurs when $\sigma^2 < 1$, meaning that the variance is less than the mean. Both situations, however, do not affect the consistency of $\widehat{\boldsymbol{\beta}}_{FEP}$. In addition, there is no restriction on arbitrary time dependence of the dependent variable within cross sections, or dependence between y_{it} and y_{ir}, $t \neq r$.

Furthermore, one can construct a variance estimator, $\hat{V}$, that is robust against heteroskedasticity, autocorrelation and misspecification of the Poisson distribution. Thus, the assumption of independence across observations within a panel need not hold; arbitrary time dependence within a panel is allowed. Also, this allows for deviations from the Poisson distribution. Construction of the estimator $\hat{V}$ begins by observing that the FE Poisson estimator, $\widehat{\boldsymbol{\beta}}_{FEP}$, is equivalent to the GMM estimator, $\widehat{\boldsymbol{\beta}}_{GMM}$. The GMM estimator is derived in the following way:

Given the following population moment condition:

$$\mathrm{E}\left[{\partial l_i(\boldsymbol{\beta})}/{\partial \boldsymbol{\beta}}\right] = \mathbf{0}$$

And the following sample moment:

$$g_N = N^{-1}\sum_{i=1}^{N}\left({\partial l_i(\boldsymbol{\beta})}/{\partial \boldsymbol{\beta}}\right)$$

The GMM estimator, $\widehat{\boldsymbol{\beta}}_{GMM}$, will be chosen to solve the following:

$$g_N = N^{-1}\sum_{i=1}^{N}\left({\partial l_i(\widehat{\boldsymbol{\beta}}_{GMM})}/{\partial \widehat{\boldsymbol{\beta}}_{GMM}}\right) = 0$$

Recall that the FE Poisson estimator, $\widehat{\boldsymbol{\beta}}_{FEP}$, is chosen to solve:

$$\sum_{i=1}^{N}\left({\partial l_i(\widehat{\boldsymbol{\beta}}_{FEP})}/{\partial \widehat{\boldsymbol{\beta}}_{FEP}}\right) = 0$$

Because minimization of $N^{-1}\sum_{i=1}^{N}\left({\partial l_i(\boldsymbol{\beta})}/{\partial \boldsymbol{\beta}}\right)$ is equivalent to minimization of $\sum_{i=1}^{N}\left({\partial l_i(\boldsymbol{\beta})}/{\partial \boldsymbol{\beta}}\right)$, the estimators are identified, i.e., $\widehat{\boldsymbol{\beta}}_{FEP} \equiv \widehat{\boldsymbol{\beta}}_{GMM}$. While both of these estimators are consistent, $\widehat{\boldsymbol{\beta}}_{FEP}$ is only efficient if the rather stringent Poisson variance assumption holds: $\mathrm{Var}(y|\mathbf{x}) = \mathrm{E}(y|\mathbf{x})$. Thus, in order to find a covariance estimator that is robust against this assumption, we must use the GMM covariance estimator using the following framework[5]:

$$\hat{V} = \widehat{\mathbf{A}}^{-1}\widehat{\mathbf{B}}\widehat{\mathbf{A}}^{-1}$$

where $\widehat{\mathbf{A}}^{-1}$is the conventional MLE variance estimator:

[5] This derivation is based on that of StataCorp (2009) and Wooldridge (2002).

$$\widehat{\mathbf{A}}^{-1} = \left({\partial g_N}/{\partial \widehat{\boldsymbol{\beta}}_{GMM}} \right)^{-1} = \sum_{i=1}^{N} \left({\partial^2 l_i(\widehat{\boldsymbol{\beta}}_{GMM})}/{\partial \widehat{\boldsymbol{\beta}}_{GMM}{}^2} \right)^{-1}$$

and **B** is estimated as:

$$\widehat{\mathbf{B}} = \sum_{i=1}^{N} \left({\partial l_i(\widehat{\boldsymbol{\beta}}_{GMM})}/{\partial \widehat{\boldsymbol{\beta}}_{GMM}} \right) \left({\partial l_i(\widehat{\boldsymbol{\beta}}_{GMM})}/{\partial \widehat{\boldsymbol{\beta}}_{GMM}} \right)'$$

where:

$$ {\partial l_i(\widehat{\boldsymbol{\beta}}_{GMM})}/{\partial \widehat{\boldsymbol{\beta}}_{GMM}} = \sum_{t=1}^{T} y_{it} \left[\left(\frac{\partial p_{it}}{\partial \widehat{\boldsymbol{\beta}}_{GMM}} \right)' / p_{it} \right]$$

Thus, the GMM estimator of $\boldsymbol{\beta}$, $\widehat{\boldsymbol{\beta}}_{GMM}$, has the following limiting distribution:

$\sqrt{\mathrm{N}}\left(\widehat{\boldsymbol{\beta}}_{GMM} - \boldsymbol{\beta}\right) \overrightarrow{d}\ N\left(0, \widehat{\mathbf{A}}^{-1}\widehat{\mathbf{B}}\widehat{\mathbf{A}}^{-1}\right)$. My dataset contains over 86,500 observations; I contend this is large enough to justify using the limiting distribution.

Chapter 5. The Data

5.1 Original Variables

The following model was used for the original FE regression[6]:

$$citation_{ijt} = \ddot{immigration}_{ijt}\beta_1 + \ddot{GDP}_{ijt}\beta_2 + \ddot{trade}_{ijt}\beta_3 + \ddot{patent\,stock}_{ijt}\beta_4 + \ddot{education}_{ijt}\beta_5 + u_{ijt}$$

Where $i = 1, 2, \dots, 1284\ US\ patents$

$j = 1, 2, \dots, 13\ Countries$

and $t = 1995, 1996, \dots, 2010\ Years.$

The variables in this model are defined below:

Citation = Number of times a US patent is cited by an inventor at time t

Immigration = Sum of total immigration to US for five years prior to time t

GDP = Gross domestic product in current US dollars at time t

Trade = Imports of US goods plus exports to US in millions of current US dollars at time t

Patent Stock = Sum of total patents and patent applications, all classes, at time t

Education = School life expectancy (years), primary to secondary, at time t

[6] For a list of variable definitions and sources, see Appendix C.

The following model was used for the original FE Poisson regressions:

$$citation_{ijt} = \exp\,(immigration_{ijt}\beta_1 + GDP_{ijt}\beta_2 + trade_{ijt}\beta_3 + patent\ stock_{ijt}\beta_4 + education_{ijt}\beta_5 + a_i) + u_{ijt}$$

where $c_i = \exp(a_i)\,,\ i = 1,\ 2,\ \ldots,\ 467\ US\ patents$

$j = \ 1, 2, \ldots, 13\ Countries$

and $t = \ 1995, 1996, \ldots, 2010\ Years.$

5.2 New Variables

Running the above regressions resulted in coefficients for immigration that were not statistically significant. Thus, it became necessary to revisit the model to add new variables and to redefine some variables that were included in the original regression. The following model was used for the new FE regression:

$$citation_{ijt} = \ddot{immigration}_{ijt}\beta_1 + \ddot{GDP}_{ijt}\beta_2 + \ddot{FDI}_{ijt}\beta_3 + \ddot{trade}_{ijt}\beta_4 + \ddot{patent\ stock\ ag}_{ijt}\beta_5 + \ddot{HDI}_{ijt}\beta_6 + u_{ijt}$$

Where $i = 1, 2, \ldots, 1284\ US\ patents$

$j = 1, 2, \ldots, 13\ Countries$

and $t = 1995, 1996, \ldots, 2010\ Years$

Immigration was not included in the revised regression because the human development index (HDI) includes an education variable. The variables in this model previously not defined or redefined are defined below:

<u>Immigration</u> = sum of total *employment-based* immigration to US for five years prior to time t

<u>FDI</u> = Inward foreign direct investment stock in millions of current US dollars at time t

<u>Patent Stock Ag</u> = Sum of total *agricultural* patents and patent applications at time t

<u>HDI</u> = Human development index at time t

The following model was used for the new FE Poisson regressions:

$$citation_{ijt} = \exp\,(immigration_{ijt}\beta_1 + GDP_{ijt}\beta_2 + FDI_{ijt}\beta_3 + trade_{ijt}\beta_4$$
$$+ patent\ stock\ ag_{ijt}\beta_5 + HDI_{ijt}\beta_6 + a_i) + u_{ijt}$$

where $c_i = \exp(a_i)$, $i = 1,\ 2,\ \ldots,\ 468\ US\ patents$

$j = 1, 2, \ldots, 13\ Countries$

and $t = 1995, 1996, \ldots, 2010\ Years$

Several of the variables in the revised model require more explanation than a simple definition. These explanations are included in the following sections.

5.3 Citation

The dataset originally included 1284 US patents, but 816 were thrown out in the FE Poisson regression because they did not have any forward citations from any country in any year. The time period was 16 years, from 1995-2010, and there were 13 sample countries. One may note that the time period starts *before* that of the sample of US patents. This is because different dates were used to define the US patents and the forward citations. In addition, the fact that the forward citation cited the US patents reveals that the knowledge or technology did indeed flow from the US to the foreign inventor. Please see the appendix for more details.

In the end, the set contained a total of 797 forward citations from inventors residing in member countries. Below, the forward citations are broken down by sample country:

Table 1. Forward Citations by Inventor Country

#	Code	Country	Citations	% Total
1	AU	Australia	69	8.68
2	BE	Belgium	24	3.02
3	CH	Switzerland	15	1.89
4	DE	Germany	159	20
5	DK	Denmark	15	1.76
6	ES	Spain	57	7.17
7	FI	Finland	11	1.38
8	FR	France	121	15.09
9	GB	Great Britain	148	18.62
10	GR	Greece	4	0.50
11	JP	Japan	84	10.57
12	NL	Netherlands	87	10.94
13	TR	Turkey	3	0.38

It is important to note that while these inventors currently *reside* in the sample countries, they may be applying for patents elsewhere. Below the data are broken down by location of patent application:

Table 2. Forward Citations, EPO Patents/Patent Applications by Inventor Country

#	Code	Country	Citations
1	AU	Australia	3
2	BE	Belgium	1
3	CH	Switzerland	1
4	DE	Germany	21
5	DK	Denmark	4
6	ES	Spain	5
7	FI	Finland	2
8	FR	France	13
9	GB	Great Britain	5
10	GR	Greece	0
11	JP	Japan	10
12	NL	Netherlands	8
13	TR	Turkey	0

The above table shows the amount of inventors from each sample country that applied for or was granted patents through the EPO.

Table 3. Forward Citations, USPTO Patents/Patent Applications by Inventor Country

#	Code	Country	Citations
1	AU	Australia	34
2	BE	Belgium	15
3	CH	Switzerland	7
4	DE	Germany	68
5	DK	Denmark	4
6	ES	Spain	8
7	FI	Finland	6
8	FR	France	32
9	GB	Great Britain	38
10	GR	Greece	4
11	JP	Japan	71
12	NL	Netherlands	26
13	TR	Turkey	0

The above table shows the amount of inventors from each sample country who applied for or were granted patents through the USPTO.

Table 4. Forward Citations, WIPO Patents/Patent Applications by Inventor Country

#	Code	Country	Citations
1	AU	Australia	32
2	BE	Belgium	4
3	CH	Switzerland	7
4	DE	Germany	31
5	DK	Denmark	5
6	ES	Spain	19
7	FI	Finland	3
8	FR	France	25
9	GB	Great Britain	27
10	GR	Greece	0
11	JP	Japan	2
12	NL	Netherlands	31
13	TR	Turkey	3

The above table shows the amount of inventors from each sample country who applied for or were granted patents through the WIPO. Finally, the table below summarizes the amount of forward citations in each location, regardless of residence of inventor:

Table 5. Forward Citations by Country/Patent Office

Code	Country/Patent Office	Citations	%Total
AU	Australia	0	0
BE	Belgium	1	0.13
CH	Switzerland	0	0
DE	Germany	42	5.27
DK	Denmark	0	0
EP	EPO	73	9.16
ES	Spain	25	3.14
FI	Finland	0	0
FR	France	50	6.27
GB	Great Britain	81	10.16
GR	Greece	0	0
JP	Japan	1	0.13
NL	Netherlands	24	3.01
TR	Turkey	0	0
US	USPTO	311	39.02
WO	WIPO	189	23.71

As you can see from the table above, the majority of patent activity is taking place in the US and the two regional patent offices contained in the sample. Together, these three patent offices combine for nearly 72% of the total patent activity.

5.4 Immigration[7]

Currently, immigration into the United States is reported by the Department of Homeland Security in annual yearbooks of immigration statistics. Before the Department of Homeland Security was created, the United States Department of Justice published these annual yearbooks. Because of the change in department oversight, there are some inconsistencies in the types of statistics reported annually. Thus, creating a consistent dataset over my data range has posed numerous difficulties. Immigrants are defined by US immigration law as "…persons lawfully admitted for permanent residence in the United States." Total immigration by country of last residence is one statistic that has remained constant, and is the first definition of immigration I use. One could easily hypothesize, however, that total migration is too broad of a measure when trying to account for those immigrants who will increase the stock of knowledge in the US and facilitate knowledge flows to source countries. This hypothesis was somewhat supported when the first regression models, using the original data, failed to report a statistically significant coefficient on immigration.

Though statistics regarding specific employment exist in the immigration yearbooks, they do not indicate from which countries these migrants are emigrating. The yearbooks do, however, report statistics on "preference immigrants" each year. Among

---

[7] Information contained in this section was taken from the 1997 Statistical Yearbook of the Immigration and Naturalization Service.

these are “employment-based preference immigrants”, which consist of the following groups of immigrants:

> “…priority workers; professionals with advanced degrees or aliens of exceptional ability; skilled workers, professionals (without advanced degrees), and needed unskilled workers; special immigrants (e.g., ministers, religious workers, and employees of the U.S. government abroad); and employment creation immigrants or ‘investors’.”

Thus, this group will include those immigrants allowed into the US for specific, industry-based purposes. Theoretically, this would be the group most likely to contribute to technological advance via patent activity. It is important to note here that spouses and children are also included in the employment preference. So, while, it would still be important to further narrow the definition in the future, I believe this measure of immigration is more accurate than total immigration and does, in fact, provide more statistically significant regression results.

5.5 Patent Stock

The original patent stock variable contained all patents granted and patent applications submitted by each of the sample countries. Like citations, this data was found using the Espacenet search engine. The patent stock for each country for each year was found by searching the publication number of the patent/patent application, which includes the country code, and the publication date, which includes the date of the patent/patent application. For example, to find all patents granted and patent applications submitted by Australia in 1995, the following advanced search criteria is used:

Publication Number: AU

Publication Date: 1995

The following table summarizes the data for all of the years contained in the dataset:

Table 6. Patent Stock by Country

#	Code	Country	Patents	% Total
1	AU	Australia	821149	8.61
2	BE	Belgium	11646	0.12
3	CH	Switzerland	16458	0.17
4	DE	Germany	1466191	15.38
5	DK	Denmark	107059	1.12
6	ES	Spain	331905	3.48
7	FI	Finland	43425	0.46
8	FR	France	239717	2.51
9	GB	Great Britain	191241	2.01
10	GR	Greece	27189	0.29
11	JP	Japan	6194203	64.95
12	NL	Netherlands	43526	0.46
13	TR	Turkey	42464	0.45

Total: 9536173

Because the above table summarizes the *total* patent stock of each country, it consists of many patents not related to plant husbandry, the class of US patents used to create my original sample of US patents. Thus, a more appropriate measure of patent stock for each country would be one analogous to plant husbandry. However, the US patents were found using the USPTO Patent Classification System. Espacenet uses the International Patent Classification (IPC) system. No exact match to plant husbandry exists in the IPC system. It does, however, have a section, Section A – Human Necessities, which contains the subclass A01 – Agriculture, Forestry, Animal Husbandry, Hunting, Trapping, and Fishing. Though this is not an exact match, it drastically narrows the definition of patent stock, making it much more relevant to the model. Note here that I use the term "agricultural patents" to define this group of patent stocks. For the search, to find all agricultural patents granted and patent applications submitted by Australia in 1995, the following advanced search criteria is used:

Publication Number: AU

Publication Date: 1995

International Patent Classification (IPC): A01

The following table summarizes the data for all of the years contained in the dataset:

Table 7. Agricultural Patent Stock by Country

#	Code	Country	Patents	% Total
1	AU	Australia	38802	16.19
2	BE	Belgium	478	0.20
3	CH	Switzerland	356	0.15
4	DE	Germany	33748	14.08
5	DK	Denmark	6544	2.73
6	ES	Spain	12318	5.14
7	FI	Finland	1590	0.66
8	FR	France	7344	3.06
9	GB	Great Britain	4809	2.01
10	GR	Greece	1641	0.68
11	JP	Japan	125906	52.53
12	NL	Netherlands	3831	1.60
13	TR	Turkey	2295	0.96

Total: 239662

5.6 Human Development Index

Each year, the United Nations Development Program publishes a Human Development Report that contains, among other things, a Human Development Index (HDI) for each country. In general, the HDI is defined in the Human Development Report (2011) as "…a summary measure of human development. It measures the average achievements in a country in three basic dimensions of human development: a long and healthy life, access to knowledge and a decent standard of living." This index, however, has changed over the years. Because the compilation of the HDI has varied over the range of dates in my dataset, it is possible that these changes may affect the results of the regressions. Thus, I have added a dummy variable to capture these changes. The dummies for the different indexes are as follows:

<u>Index 1[8]:</u>

Years: 2008-2010

The HDI is the geometric mean of the three dimension indices:

$$HDI = I_{Life}^{1/3} \cdot I_{Education}^{1/3} \cdot I_{Income}^{1/3}$$

Where:

$$I_{Life} = Life\ Expectancy\ Index$$

$$= \frac{actual\ value\ of\ life\ expectancy - 20}{maximum\ value\ of\ life\ expectancy\ observed - 20}$$

[8] Compilation of this index was acquired from the Human Development Report 2011.

$I_{Education} = Education\ Index$

$$= \frac{\sqrt{I_{Mean\ Years\ of\ Schooling} \cdot I_{Expected\ Years\ of\ Schooling}} - 0}{maximum\ value\ of\ Education\ Index\ Observed - 0}$$

$I_{Mean\ Years\ of\ Schooling} = \ Mean\ Years\ of\ Schooling\ Index$

$$= \frac{actual\ mean\ years\ of\ schooling - 0}{maximum\ value\ of\ mean\ years\ of\ schooling\ observed - 0}$$

$I_{Expected\ Years\ of\ Schooling} = Expected\ Years\ of\ Schooling\ Index$

$$= \frac{actual\ expected\ years\ of\ schooling - 0}{maximum\ value\ of\ expected\ years\ of\ schooling\ observed - 0}$$

$$I_{Income} = \ Income\ Index = \frac{\ln(actual\ per\ capita\ income) - \ln{(100)}}{\ln(maximum\ per\ capita\ income\ observed) - \ln{(100)}}$$

<u>Index 2[9]:</u>

Years: 1997, 1998,1999, 2000, 2001-2004, 2005, 2006, 2007

The HDI is the simple average of the three dimension indices:

$$\left(\frac{1}{3}\right) I_{Life} + \left(\frac{1}{3}\right) I_{Education} + \left(\frac{1}{3}\right) I_{GDP}$$

Where:

$I_{Life} = Life\ Expectancy\ Index$

$$= \frac{actual\ value\ of\ life\ expectancy - 25}{maximum\ value\ of\ life\ expectancy\ observed - 25}$$

$$I_{Education} = Education\ Index = \left(\frac{2}{3}\right) I_{Adult\ Literacy} + (\frac{1}{3}) I_{Gross\ Enrollment}$$

[9] Compilation of this index was acquired from the Human Development Report 2006.

$$I_{Adult\ Literacy} = Adult\ Literacy\ Index = \frac{actual\ adult\ literacy\ rate\ - 0}{100 - 0}$$

$$I_{Gross\ Enrollment} = Gross\ Enrollment\ Index = \frac{actual\ gross\ enrollment\ ratio - 0}{100 - 0}$$

$$I_{GDP} = GDP\ Index = \frac{\ln(actual\ per\ capita\ GDP) - \ln(100)}{\ln(40{,}000) - \ln(100)}$$

<u>Index 3[10]:</u>

Years: 1995

The HDI is the simple average of the three dimension indices:

$$\left(\frac{1}{3}\right) I_{Life} + \left(\frac{1}{3}\right) I_{Education} + \left(\frac{1}{3}\right) I_{GDP}$$

Where:

$$I_{Life} = Life\ Expectancy\ Index$$

$$= \frac{actual\ value\ of\ life\ expectancy - 25}{maximum\ value\ of\ life\ expectancy\ observed - 25}$$

$$I_{Education} = Education\ Index = \left(\frac{2}{3}\right) I_{Adult\ Literacy} + (\frac{1}{3}) I_{Gross\ Enrollment}$$

$$I_{Adult\ Literacy} = Adult\ Literacy\ Index = \frac{actual\ adult\ literacy\ rate\ - 0}{100 - 0}$$

$$I_{Gross\ Enrollment} = Gross\ Enrollment\ Index = \frac{actual\ gross\ enrollment\ ratio - 0}{100 - 0}$$

[10] Compilation of this index was acquired from the Human Development Report 1998.

Construction of the GDP index is based on Atkinson's formula for the utility of income:

Let $y^* = \$5{,}999$ GDP per capita = threshold level of income, and y = actual GDP per capita.

The adjusted real GDP per capita, W(y) is calculated as:

$$
\begin{aligned}
W(y) &= y^* \text{ for } 0 < y < y^* \\
&= y^* + 2[(y-y^*)^{1/2}] \text{ for } y^* < y < 2y^* \\
&= y^* + 2(y^{*1/2}) + 3[(y-y^*)^{1/3}] \text{ for } 2y^* < y < 3y^*
\end{aligned}
$$

$$I_{GDP} = GDP\ Index = \frac{W(y) - 100}{W(y_{maximum\ observed}) - 100}$$

Chapter 6. Results

Before performing any regression analysis, it is helpful to make predictions regarding the signs of the variable coefficients in the model. Below is a table summarizing my predictions for the signs of all (original, new, and redefined) variable coefficients.

Table 8. Predicted Signs of Variable Coefficients

Variable	Predicted Sign of Coefficient
GDP	+
Trade	+
FDI	+
Patent Stock	+
Patent Stock Ag	+
Education	+
HDI	+
Immigration	+
Employment-based Immigration	+

When predicting the signs, it is necessary to remember the implication: a positive coefficient for any given variable means that an increase in the value of that variable will increase the amount of foreign inventors citing US patents, all else equal. The sign of GDP is expected to be positive; prior literature has shown that, holding other variables constant, patents issued are positively correlated with GDP. Thus, a country with a higher GDP is issuing more patents, citing more patents in general, and most likely citing more US patents as well. Likewise, the positive correlation between trade and patent citations was shown empirically by Sjöholm in the article on which I loosely base my estimation technique. The argument of a positive correlation between patent stock and citations is parallel to that of GDP's correlation with citations; more patent activity leads to more patent citations in general, including citations of US patents.

However, as discussed earlier, one would expect including only agricultural patents and patent applications to be a better measure of patent activity related to the US patents whose forward citations are being sought and documented, simply because the US patents are agricultural patents. It is fairly simple to argue the positive correlation between education and patent citations; higher education leads to a more skilled labor force, which would then be more likely to create technological advance via patents.

Similarly, HDI is an index comprised of educational variables mainly, and so too is expected to have a positive coefficient. Finally, of course, all immigration variables are expected to be positive as theorized in this paper's hypothesis. However, employment-based immigration is expected to be a better measure of those immigrants contributing to the stock of knowledge and technology in the US, thus increasing the probability of sending that knowledge back to the source countries.

6.1 Original Variables

Table 9. FE OLS, Original Variables

Variable	Coefficient (Standard Error)	t-statistic	95% Confidence Interval	
GDP	6.46e-16 (3.68e-16)	1.75	-7.63e-17	1.37e-15
Trade	5.19e-08 (1.03e-08)	5.05**	3.17e-08	7.21e-08
Patent Stock	-2.17e-08 (3.45e-09)	-6.30**	-2.85e-08	-1.49e-08
Education	.0005314 (.0000717)	7.41**	.0003907	.000672
Immigration	1.38e-08 (1.21e-08)	1.14	-1.00e-08	3.76e-08
Note: **Significant at the 5 percent level. $R^2 = 0.0015$				

The above table summarizes the results obtained from running pooled OLS on the original FE model. As mentioned before, the purpose of running this regression was to have a basis for the coefficients when running the more appropriate model, the FE Poisson regression model. Thus, what is important to note is the sign of the coefficients. As predicted, GDP, trade, education and immigration all have positive coefficients, which means that increasing these will increase the amount of patents cited by inventors from one of the sample countries.

Somewhat interestingly, the coefficient on patent stock is negative. This could be counterintuitive, as one might hypothesize that countries with larger patent stocks would cite more patents, including those from the US. However, if we look at the patent stock of each country for all patents during the same time period, and each country's percentage of the total, we see that there is an obvious difference:

Table 10. Patent Stock by Country

#	Code	Country	% Citations	% Patents
1	AU	Australia	8.68	8.61
2	BE	Belgium	3.02	0.12
3	CH	Switzerland	1.89	0.17
4	DE	Germany	20	15.38
5	DK	Denmark	1.76	1.12
6	ES	Spain	7.17	3.48
7	FI	Finland	1.38	0.46
8	FR	France	15.09	2.51
9	GB	Great Britain	18.62	2.01
10	GR	Greece	0.50	0.29
11	JP	Japan	10.57	64.95
12	NL	Netherlands	10.94	0.46
13	TR	Turkey	0.38	0.45

As the above table reveals, Japan constitutes nearly 65 % of the patents, but less than 11% of the forward citations. This is most likely the reason for the negative coefficient, and can possibly attributed to the fact that Japan has "less in common" with the US than the other countries. For example, cultural differences exist that could hinder communication or other factors, thus decreasing the amount of technology flowing from the US to Japan and thus forward citations.

Finally, the last thing to note is the very low R-squared value. This is not surprising, however, as the data is much better characterized by a Poisson distribution since it is strongly skewed to the right. Thus, we would expect a goodness of fit test to be "bad".

Table 11. FE Poisson, Original Variables

Variable	Coefficient (Standard Error)	z-statistic	95% Confidence Interval	
GDP	1.99e-13 (9.11e-14)	2.19**	2.08e-14	3.78e-13
Trade	.0000139 (2.54e-06)	5.45**	8.88e-06	.0000189
Patent Stock	-5.23e-06 (6.49e-07)	-8.05**	-6.50e-06	-3.96e-06
Education	.27509 (.0303498)	9.06**	.2156054	.3345746
Immigration	6.97e-07 (2.19e-06)	0.32	-3.59e-06	4.99e-06
Note: **Significant at the 5 percent level.				

The above table summarizes the results obtained from performing conditional MLE on the original FE Poisson regression model. Because this is not a linear regression model, the interpretation of the coefficients is not as straightforward as the interpretation of the coefficients from the OLS model:

$$\mathrm{E}(y|\mathbf{x},\ c) = \mathrm{c}\exp(\mathbf{x}\boldsymbol{\beta}), \text{ thus } \frac{\partial \mathrm{E}(y|\mathbf{x}, c)}{\partial x_j} = \mathrm{c}\exp(\mathbf{x}\boldsymbol{\beta})\,\beta_j = \mathrm{E}(y|\mathbf{x},\ c)\beta_j$$

$$\text{So } \beta_j = \frac{\partial \mathrm{E}(y|\mathbf{x}, c)}{\partial x_j} \cdot \frac{1}{\mathrm{E}(y|\mathbf{x}, c)} = \frac{\partial \log\,[\mathrm{E}(y|\mathbf{x})]}{\partial x_j}$$

If x_j changes by 1 unit, $\partial x_j = 1$ and:

$$\beta_j = \frac{\partial \log\,[\mathrm{E}(y|\mathbf{x},\ c)]}{1} = \partial \log[\mathrm{E}(y|\mathbf{x},\ c)]$$

Thus, the coefficients can be interpreted in the following manner: for a one unit change in the independent variable, that variable’s coefficient is equal to the change in the difference in the logs of the predicted amount of forward citations, holding all other independent variables constant. For example, the above table reveals that, all else equal, if GDP of a sample country increases by a dollar, the difference in the logs of expected forward citations from that country in a given year will increase by 1.99e-13. Though this number is very small due to the large amount of zero counts of forward citations in the dataset, it is nonetheless statistically significantly greater than zero. One will note that the signs of the coefficients from the FE Poisson regression are all the same as those from the results of the FE model in Table 3. In addition, all variables' coefficients except that of lagged immigration are statistically significant at the five percent level. Thus, though the effect of immigration on the expected amount of forward citations from a given country at a given time is positive as expected, it is not statistically significant.

6.2 New Variables

Though the original model contained vital regressors, numerous, important variables were not included. As discussed earlier, the patent stock and immigration variables were redefined, and the variables of FDI and HDI were added. The table below summarizes the results obtained from running pooled OLS on the new FE model.

Table 12. Fixed Effects OLS, New Variables, HDI Index

Variable	**Coefficient (Standard Error)**	**t-statistic**	**95% Confidence Interval**	
GDP	-2.71e-16 (3.14e-16)	-0.86	- 8.87e-16	3.46e-16
FDI	4.74e-09 (6.97e-10)	6.81**	3.38e-09	6.11e-09
Trade	1.08e-08 (9.38e-09)	1.16	-7.56e-09	2.92e-08
Patent Stock Ag	2.99e-07 (1.53e-07)	1.95	-1.39e-09	5.99e-07
HDI Index 1	.0024989 (.0011694)	2.14**	.0002047	.004793
HDI Index 2	.0061635 (.0011092)	5.56**	.0039874	.0083396
HDI Index 3	.0037433 (.0010783)	3.47**	.0016279	.0058587
Immigration	6.29e-08 (2.41e-08)	2.61**	1.57e-08	1.10e-07
Constant	-.0038226 (.0009561)	-4.00**	-.0056983	-.0019469
Note: **Significant at the 5 percent level. $R^2 = 0.0018$				

Again, the purpose of running this regression was to have a basis for the coefficients when running the more appropriate model, the FE Poisson regression model.

Thus, we merely note the sign of the coefficients here. After reformulating the model, the coefficient on GDP became negative, which is opposite of the expected sign, but the value is statistically insignificant. Likewise, the coefficient on patent stock became positive, as expected, though it too is statistically insignificant. As predicted, the coefficients on FDI and HDI (all three indexes) are both positive. The R-squared value of 0.0018 is very small; but again, not surprising, as OLS does not accurately predict the dependent variable. In fact, I have reported the constant term in the above table to indicate that it is indeed theoretically possible to get values of $\mathbf{x}$ where $\mathbf{x}\hat{\boldsymbol{\beta}}_{FE} < 0$. The most important note here is that the coefficient on immigration is still positive, and, though this is not the appropriate model, statistically significant.

Table 13. Fixed Effects Poisson, New Variables, HDI Index

Variable	Coefficient (Standard Error)	z-statistic	95% Confidence Interval	
GDP	-4.65e-14 (1.03e-13)	-0.45	-2.48e-13	1.55e-13
FDI	1.06e-06 (1.63e-07)	6.50**	7.41e-07	1.38e-06
Trade	3.57e-06 (2.60e-06)	1.37	-1.53e-06	8.67e-06
Patent Stock Ag	.0000603 (.0000397)	1.52	-.0000175	.000138
HDI Index 1	3.96721 (.8027314)	4.94**	2.393885	5.540534
HDI Index 2	5.000825 (.7436038)	6.73**	3.543388	6.458261
HDI Index 3	3.885115 (.7870518)	4.94**	2.342522	5.427709
Immigration	.0000106 (4.29e-06)	2.46**	2.14e-06	.000019
Note: **Significant at the 5 percent level.				

The above table summarizes the results obtained from performing conditional MLE on the new FE Poisson regression model. The signs of the variable coefficients are all the same as the OLS model. Here, however, it is more appropriate to note the significance as well as the signs of the variable coefficients. The sign of GDP is not as predicted, but is statistically insignificant. Both FDI and HDI are positive, as expected, and statistically significant. Though trade and patent stock both have positive coefficients like expected, they are not statistically significant. Finally, and most importantly, the coefficient on immigration is now positive and statistically significant. As seen from the results in the above table, an increase of 100 immigrants into the US from country j in year t is associated with a .1% increase in the number of inventors in country j in year t who cite a US patent. Thus, after adding new relevant variables and redefining variables, the model now reveals that immigration does have a positive and statistically significant effect on the amount of foreign inventors that cite US patents. In other words, in this formulation, knowledge flows have been found to be positively correlated with migration flows.

Chapter 7. Sensitivity Analysis

The empirical literature in economics that exists investigating the correlation between two variables is vast. However, Cooley and LeRoy (1981) note that economic theory "…does not generate a complete specification of which variables are to be held constant when statistical tests are performed on the relation between the dependent variable and the independent variables of primary interest." Because of this, many of the empirical studies only use very specific models with a relatively small number of explanatory variables in order to report a statistically significant relationship between two variables of interest. As a result, the majority of conclusions drawn in the literature are fragile; they depend on the conditioning set of information in the regression model. Two sensitivity analyses are performed below in hopes of providing "full disclosure" and reporting robust results.

7.1 Immigration Variables

The original immigration variable, as discussed earlier, consisted of the sum of all immigrants into the US from the five years prior to time t. The new immigration variable, also as discussed earlier, consisted of the sum of only employment-based immigrants into the US from the five years prior to time t. In order to look more closely at specific years, I have rerun the FE Poisson regression model, using the new variables, but varying the immigration variable. The results are reported in the table below[11]:

[11] Complete tables of results from each regression are reported in Appendix E.

Table 14. Sensitivity Analysis, Immigration Variables

Variable	Coefficient (Standard Error)	z-statistic	Sign	Significant
Original Immigration Total	.0000114 (2.54e-06)	4.48**	+	Yes
Original Immigration Lag 1	.0000504 (.0000115)	4.38**	+	Yes
Original Immigration Lag 2	.0000366 (.0000117)	3.13**	+	Yes
Original Immigration Lag 3	. 0000508 (.000011)	4.62**	+	Yes
Original Immigration Lag 4	.0000535 (9.71e-06)	5.51**	+	Yes
Original Immigration Lag 5	.0000301 (.0000105)	2.88**	+	Yes
New Immigration Total	.0000106 (4.29e-06)	2.46**	+	Yes
New Immigration Lag 1	.0000435 (.0000194)	2.24**	+	Yes
New Immigration Lag 2	-5.56e-07 (.8080153)	-0.03	-	No
New Immigration Lag 3	.0000561 (.0000217)	2.58**	+	Yes
New Immigration Lag 4	.0000635 (.0000189)	3.36**	+	Yes
New Immigration Lag 5	-.0000225 (.0000217)	-1.04	-	No
Note: **Significant at the 5 percent level.				

As you can see in the table above, the immigration variable is quite robust. Only employment-based immigration from two and five years prior to time t are negative, and these results are statistically insignificant. All other immigration variables are positive and statistically significant.

7.2 Independent Variables

A different type of sensitivity analysis involves changing the independent variables in the model and investing the results. In their article "Reporting the Fragility of Regression Estimates", authors Leamer and Leonard argue that no model should be taken as given. That is, the advance of econometric technology has allowed economic professionals to make many conflicting inferences drawn from the same set of data. Leamer and Leonard propose an alternative econometric technology that allows researchers to summarize the entire range of inferences implied by a whole family of alternative models using a given data set. Very simply, it is a sensitivity analysis that consists of systematically changing the parameterization of the model and reporting the results. They conduct this analysis by imposing various combinations of exclusion restrictions around one variable of interest and observe whether the coefficient on the variable of interest remain statistically significant and of the same sign. This analysis allows a reporting of results that is much more informative than the common reporting of results in the literature.

For the following sensitivity analysis, I follow one similar to that of Leamer and Leonard; one proposed by Levine and Renelt in their article "A Sensitivity Analysis of Cross-Country Growth Regressions". Like Leamer and Leonard, Levine and Renelt agree that coefficient estimates on variables of interest depend vitally on the conditioning set of information. To perform their sensitivity analysis, Levine and Renelt use data regarding the long-run growth rates and a variety of regressors linked to it in the literature. They then run numerous regressions with one chosen variable of interest, a set of variables always included, and vary another set of variables that varies for each regression. They find almost all variables of interest fragile, meaning they do not remain the same sign and statistically significant over the range of regressions.

In my sensitivity analysis, I rerun the FE Poisson regression model, keeping the sum of employment-based immigration for the five years prior to time t as the constant immigration variable, and varying the new independent variables in sets of three. There are a total of $\binom{5}{3} = 10$ regression models. The results are reported below (Note that I first report a regression consisting of immigration as the only independent variable)[12]:

[12] Complete tables of results from each regression are reported in Appendix E.

Table 15. Sensitivity Analysis, Independent Variables

Regression	Variables In Regression	Coefficient on Immigration (Standard Error)	z-statistic	Sign	Significant
0	None	.0000316 (2.23e-06)	14.15**	+	Yes
1	GDP, FDI Trade	5.92e-06 (3.73e-06)	1.59	+	No
2	GDP, FDI Patent Stock	.0000153 (3.64e-06)	4.21**	+	Yes
3	GDP, FDI HDI	8.74e-06 (4.02e-06)	2.18**	+	Yes
4	GDP, Trade Patent Stock	.0000197 (4.06e-06)	4.86**	+	Yes
5	GDP, Trade FDI	.0000237 (3.01e-06)	7.87**	+	Yes
6	GDP, Patent Stock HDI	.0000189 (3.95e-06)	4.79**	+	Yes
7	FDI, Trade Patent Stock	.0000117 (4.05e-06)	2.89**	+	Yes
8	FDI, Trade HDI	8.60e-06 (3.95e-06)	2.18**	+	Yes
9	FDI, Patent Stock HDI	.0000145 (3.52e-06)	4.11**	+	Yes
10	Trade, Patent Stock HDI	.0000193 (4.28e-06)	4.51**	+	Yes
Note: **Significant at the 5 percent level.					

Like the sensitivity analysis using immigration variables, you can see from the table above that the immigration variable in this sensitivity analysis is also fairly robust. All regressions yield a positive immigration correlation coefficient, and all but one regression yield statistically significant results. Thus, one can conclude that there is a robustly positive correlation between citations and immigration, or between knowledge flows and migration flows.

Chapter 8. New Panel Model

In the final section of this paper, I report results from a different panel regression model. In this model, I ignore the unobservable effects of US patents and include country and time effects.

As with the original model, I first use FE pooled OLS model to analyze the data. I then use the more appropriate FE Poisson model. I use all variables from the new regression models and the HDI index dummies.

Thus, the model is as follows:

FE OLS:

$$\ddot{citation}_{it} = \ddot{immigration}_{it}\beta_1 + \ddot{GDP}_{it}\beta_2 + \ddot{FDI}_{it}\beta_3 + \ddot{trade}_{it}\beta_4 + patent\ \ddot{stock}\ ag_{it}\beta_5 + \ddot{HDI}_{it}\beta_6 + u_{it}$$

Where $Country\ i = 1, 2, \dots, 13$ and $year\ t = 1995, 1996, \dots, 2010$

FE Poisson:

$$citation_{it} = \exp\left(immigration_{it}\beta_1 + GDP_{it}\beta_2 + FDI_{it}\beta_3 + trade_{it}\beta_4 + patent\ stock\ ag_{it}\beta_5 + HDI_{it}\beta_6 + a_i\right) + u_{it}$$

Where $c_i = \exp(a_i)$, $Country\ i = 1, 2, \dots, 13$ and $year\ t = 1995, 1996, \dots, 2010$

The results of the regressions are listed in the table below (Note that the expected signs of the coefficients remain the same as with the original models):

Table 16. FE OLS, New Panel Model

Variable	Coefficient (Standard Error)	t-statistic	95% Confidence Interval	
GDP	-4.03e-12 (2.48e-12)	-1.63	-9.43e-12	1.36e-12
FDI	5.20e-06 (3.71e-06)	1.40	-2.87e-06	.0000133
Trade	-.0000196 (.000015)	-1.31	-.0000524	.0000131
Patent Stock Ag	.0021771 (.0003155)	6.90**	.0014896	.0028646
HDI Index 1	29.54556 (15.55089)	1.90	-4.336913	63.42803
HDI Index 2	29.21662 (15.00282)	1.95	-3.471716	61.90495
HDI Index 3	24.90603 (14.70241)	1.69	-7.127763	56.93983
Immigration	-.0000424 (.0001718)	-0.25	-.0004166	.0003319
Constant	-19.87755 (14.21884)	-1.40	-50.85774	11.10264
Note: **Significant at the 5 percent level. $R^2 = 0.3210$				

Table 17. FE Poisson, New Panel Model

Variable	Coefficient (Standard Error)	z-statistic	95% Confidence Interval	
GDP	-6.46e-13 (2.45e-13)	-2.64**	-1.13e-12	-1.66e-13
FDI	5.97e-07 (5.41e-07)	1.10	-4.62e-07	1.66e-06
Trade	-3.17e-06 (5.77e-06)	-0.55	-.0000145	8.14e-06
Patent Stock Ag	.0006656 (.0000627)	10.61**	.0005426	.0007885
HDI Index 1	13.09801 (5.937998)	2.21**	1.459752	24.73628
HDI Index 2	12.72772 (5.605145)	2.27**	1.741836	23.7136
HDI Index 3	11.27374 (5.547188)	2.03**	.4014531	22.14603
Immigration	-2.24e-07 (.0000178)	-0.01	-.000035	.0000346
Note: **Significant at the 5 percent level.				

As you can see from the tables above, the results are quite different from the original regression models. In the Poisson regression, the effect GDP is negative and statistically significant; it was seldom significant in any of the regressions using the original panel model. The coefficient on immigration is negative in both regressions, but is statistically insignificant in both as well. I believe the unobservable effect of US patents was significant and correlated with the regressors, specifically immigration. Thus, removing it introduced omitted variable bias, as suggested by the above results.

Chapter 9. Conclusion

9.1 Discussion of Relevance

As one may note, the countries in my sample include some European countries plus Japan and Australia. Of course, these are developed countries. This sample is a simple result of patent data restrictions. My hypothesis, however, is intended to imply that migration to more developed countries, in this case the United States, can help the lesser developed source countries "catch-up" in an economic sense via increased technology flows and thus expedited economic growth. This result would off-set the detrimental effects of the brain drain process on source countries. Thus, it is important to consider here whether the results from my empirical analysis can be extended to less developed countries. That is, would I find the same robust correlation between migration and technology flows if my sample were extended to include lesser developed countries?

Firstly, my hypothesis builds upon the idea that migration to the US is occurring as a result of the brain drain; educated citizens of other countries are migrating to the US in search of better employment opportunities and, in some cases, higher educational attainment possibilities. While this process has been proven empirically for channels from lesser to more developed countries, the results of my analysis could only logically be extended if this were the case in my sample countries. In other words, are the countries in my sample experiences brain drain migration to the US?

In an article entitled "The European brain drain: European workers living in the US", Gilles Saint-Paul uses US and European census data to reveal that the brain drain process is indeed occurring from Western Europe to the US; Europeans living in the US are vastly outperforming both their American and European counterparts. To this point, the table below shows the percentage of the expatriate population with tertiary education versus the corresponding percentage in home country and the whole US in 1990 and 2000:

Table 18. Percentage of Population with Tertiary Education

Country	1990		2000	
	In United States	In Home Country	In United States	In Home Country
Belgium	47.6	17	59.6	26
France	42.7	14	56.1	24
Germany	34.6	17	41.9	28
Great Britain	38.9	15	49.5	25
Italy	17.1	6	25.7	13
Spain	30.6	9	44.1	21
United States	29.7	N/A	33.8	N/A

The above table reveals that Europeans living in the US are more likely than their US counterparts and approximately twice as likely as their European counterparts to have tertiary educational attainment. Furthermore, the table below shows the percentage of European expatriates with a Ph.D. as compared to the percentage of the whole US population in 1990 and 2000.

Table 19. Percentage of European Population in US with a Ph.D.

Country	1990	2000
Belgium	4.33	5.78
France	3.1	4.9
Germany	1.72	2.39
Great Britain	3.2	3.9
Italy	0.96	2.0
Spain	2.7	4.6
United States	0.82	0.98

The above table reveals that European expatriates are upwards of nearly five times more likely to hold Ph.D.s than the US as a whole.

Likewise, evidence presented by Yukiko Murakami in his article “Japan’s Brain Drain: An Analysis of Japanese Researchers Living in the United States” suggests that Japan is also suffering from the brain drain process to the US. He writes that “…a considerable number of Japanese researchers and engineers are moving overseas, primarily to the United States.” He goes on to add that “…the number of Japanese individuals living in the United States who have an undergraduate or higher level of education, and who have a degree in a field related to science or engineering is as high as 59,400[13].” Thus, the results of my analysis are still relevant in that migration leading to knowledge transfers can, in fact, help mitigate the detrimental effects associated with the brain drain process.

Secondly, my hypothesis relies on the assumption that immigrants in the US are gaining knowledge in the US and then sending that knowledge back to their locations of prior residence. Though this has been proven in my analysis for the countries in my

[13] This figure is from the National Science Board (2006).

sample, whether this is the case in lesser developed countries is less clear. One report, however, contends that if this knowledge flows via return migration, then knowledge will not flow as readily as a result of migration from lesser developed countries. According to the OECD's 2008 report "International Migration Outlook":

> "The smaller the development gap between the home and settlement country, the more likely it is that migrants will go back home; return rates to OECD countries are twice as high as those to developing countries."

Thus, immigrants to the US from developing countries are less likely than the countries in my sample to send knowledge back home via return migration.

I found no evidence to suggest, however, that migrants from developing countries are less likely to communicate with their compatriots back home. This seems especially true when looking only at employment-based labor. Though communication can be expensive, one would assume that skilled labor and employment based labor from all countries would be able to afford methods of communication.

In conclusion, it remains unclear whether one would be justified in extending these results to lesser developed countries and predicting that we would find the same positive, statistically significant correlation between migration and technology flows using a larger sample of countries. Somewhat ironically, with the achievement of technological advance in lesser developed countries, patent data may become available in the future and we will be able to conduct this empirical analysis using more comprehensive sample of countries. Only then will we have the answers we seek.

9.2 Concluding Remarks

With quick review of the literature, it becomes clear that technological progress has a distinct and real effect on immigration. As one example, a very extensive literature exists regarding the brain drain process. However, this relationship is not unidirectional. It has also been shown that migration can affect technological progress, or at least the dissemination of knowledge. If technology flows back to the source country, they need not suffer the welfare-reducing effects of brain drain. In fact, they could benefit from sending labor abroad if it meant expedited technology transfers from host countries. This process, however, remains relatively untouched in the field of economics, especially on an international scale. That is, little is known about the effects of migration on technological progress in the source country.

This paper has examined the relationship between migration and technology flows from host to source countries in a uniquely robust way. My results have provided evidence regarding the relationship between numerous variables and their effects on technological advance in source countries. More specifically, this paper has shown empirically that a positive and statistically significant relationship exists between migration flows and technology flows. This implies that migration to a host country can create knowledge or technology flows back to the host country. The majority of literature regarding brain drain migration has found that the result is a permanent increase in income and economic growth in the host country relative to that of the source country. My paper, however, finds that brain drain migration can result in benefits to the source country in the form of increased inward technology flows from abroad. Moreover, the scarce literature regarding migration and its effects on knowledge flows has been regional: regional migration data and regional patent citation information were used to perform the analyses. Virtually no literature exists regarding the relationship between international migration and international knowledge flows.

Furthermore, this relationship has been analyzed in uniquely robust way. Multiple sensitivity analyses were performed on the variable of immigration, showing the relationship between immigration and knowledge flows to lack fragility. This evidence is not currently available on an international scale in any context.

The majority of immigration literature focuses solely on the effects of immigration on labor markets. The results of this paper reach much further than the existing literature; I have found that a positive, statistically significant correlation exists between immigration and technology flows. This result has many implications for both source countries and host countries. As explained before, Solow's neoclassical growth model contends that in order to achieve an increased steady-state rate of growth, countries *must* innovate. Put simply, technological advance resulting from brain drain migration creates sustainable economic growth in source countries.

Additionally, if some return on brain drain migration exists in the form of increased inward technology flows, source countries may be more willing to accept brain drain migration from a policy standpoint, or even encourage it. If immigration is creating "brain circulation," that makes both host and source countries better off, host countries, namely the US, could benefit from relaxing its rather stringent immigration policies.

Any policy suggestions derived from the results of this paper, however, should not be wholly accepted without scrutiny. As previously mentioned, the effects of immigration reach far beyond just the labor market and changes in technology. Immigrants bring with them much more than knowledge and a supply of labor. Large-scale immigration can create cultural and political shifts that could result in unrest if allowed or encouraged with excessive haste. All effects of immigration on both source and host countries should be investigated exhaustively before any relevant policy decisions are made.

Finally, my sample includes relatively developed countries. These countries, however, also suffer from brain drain migration to the US. In order to extend the results of this paper to more developed countries, further investigation must be conducted. Because of data limitation, patent citation information is not currently available to the widespread public. However, some useful information regarding developing countries is useful in examining the relationship between migration and technology flows. Foreign direct investment (FDI) by country is data that is available for many countries, including developing countries. Likewise immigration data is available for immigrants for developing countries as well as more developed countries. We can then use this data to examine the relationship between migration to the US and inward FDI flows in developing countries. If we assume that FDI is one avenue through which technology flows from the host country to the source country, the results of this analysis will provide information regarding the relationship between migration flows and technology flows. That is, if migration to the US is found to create FDI flows into source countries, then we can argue this will create increased technology flows from the US back to source countries. Using this method, we may be able to say something about the extension of the results found in this paper to developing countries.

Though the results of this paper are quite significant, there is still much to be done in order to uncover the complete details of the correlation between technology flows and migration flows. More dependent variables can be added to the model to further isolate the effect of immigration on technology flows. I would also like to run the regression using $(\text{patent stock})^2$, implying that diminishing or increasing returns to innovation may exist.

In addition, I aim to uncover the specific ways in which migration enables knowledge to flow back to source countries. For example, does immigration to the US create incentives to send FDI to source countries, as Saxenian hypothesized? Is it the personal contact that immigrants maintain with residents remaining in source countries that helps facilitate knowledge transfer, as Agrawal, Cockburn and McHale speculated? Or, is it via return migration that knowledge flows to source countries, as concluded by Mayr and Peri?

As information technology advances, more data will become available. Thus, forward citation documentation may become available for more countries. With this information, the sample of countries in my analysis can be expanded and we can investigate whether this correlation exists in other, perhaps less developed, countries. In addition, if better migration data becomes available in the US, we can better narrow the definition of "immigration" to only include highly skilled labor. I believe this would yield even more statistically significant results. Information was not available regarding the gender or culture of inventors or the employment-based immigrants. Should this data become available, numerous interesting studies could be performed on the gender and cultural characteristics of both immigrants and inventors.

Though there is still much work to be done on this topic and this paper, some light has been shed on possible avenues through which technological advance can be achieved, technology gaps between developed and developing countries can be bridged, and sustainable long-run economic growth can be achieved.

Appendix A

Patent Sources

Espacenet

http://worldwide.espacenet.com/?locale=en_EP

Operated by the EPO, Escpacenet is an online, searchable database that contains over 70 million patent documents from 1836 to the present. See Table B1 for a list of member countries and international patent organizations.

European Patent Office (EPO)

http://www.epo.org/

Created October 7, 1977, the EPO is responsible for granting European patents and conducting search reports for patent applications submitted to various national patent offices across Europe. The EPO consists of 38 member states throughout Europe. The patents the EPO grants are not "international" patents, but rather a bundle of national patents.

World Intellectual Property Organization (WIPO)

http://www.wipo.int/patentscope/en/

Established in 1967 at the WIPO Convention, the WIPO is a United Nations agency. Its 142 Member States around the world collaborate to promote the protection of patents and intellectual property internationally. In addition, the WIPO also performs many steps of the patent application process centrally, so that these steps need not be repeated in each country that could possibly grant the patent. The first of these steps includes accepting and filing international patent applications submitted under the Patent Cooperation Treaty

(PCT). WIPO's trademarked database search engine, PATENTSCOPE, allows one to perform advanced searches of over 1.8 million of these applications.

Appendix B

EPO Database Contents

Table B1. Member Countries and International Patent Organizations in EPO's Citation Database

Country Code	Country/Organization	First Publication Date
AP	ARIPO	7/3/1985
AU	Australia	3/18/1971
BE	Belgium	12/15/1987
CH	Switzerland	6/30/1963
DE	Germany	9/18/1943
DK	Denmark	2/6/1956
EP	European Patent Office	12/20/1978
FI	Finland	12/31/1990
FR	France	8/29/1969
GB	Great Britain	1/4/1979
GR	Greece	1/19/1990
JP	Japan	11/9/1965
NL	Netherlands	2/15/1947
TR	Turkey	1/7/1987
US	United States of America	1/7/1947
WO	WIPO	10/19/1978

Appendix C

Glossary of Terms

Table C1. Variable Definitions and Data Sources

Variable	Source	Definition/Description
US Patent	US Patent and Trademark Office – Patent Full-Text and Image Database (PatFT) http://patft1.uspto.gov/netahtml/PTO/search-adv.htm	A patent from current **US patent class 47** (including all subclasses) officially issued to inventors who were US residents at the time of publication. The dataset consists of patents issued by the US Patent and Trademark Office between the dates 1/1/98 – 12/31/02 and contains 1284 patents.
US Patent Class 47 "Plant Husbandry"	US Patent and Trademark Office – Patent Full-Text and Image Database (PatFT) http://patft1.uspto.gov/netahtml/PTO/search-adv.htm	One of 1022 US patent classes, US patent class 47 is entitled "Plant Husbandry" and contains 89 subclasses. It is defined by the US Patent and Trademark as "…the parent class for apparatus and processes employed in treating the earth and its products and includes all inventions relating thereto that have not been especially provided for in other classes."
Citation, or citing document	Espacenet http://worldwide.espacenet.com/?locale=en_EP	Also called a forward citation, a patent issued to or patent application from an inventor who is a resident of one of the sample countries that has cited one of the **US patents**. There are a total of 797 citations.
Country	Espacenet http://worldwide.espacenet.com/?locale=en_EP	The country of residence of the person named in a **citing document** as the inventor. This information is provided by the person, applicant or inventor filing the form.
Priority Date	Espacenet http://worldwide.espacenet.com/?locale=en_EP	The date assigned to a patent application when it is filed. This is the earliest date associated with the foreign patent. These dates range from 1995-2010.
Publication Date	Espacenet http://worldwide.espacenet.com/?locale=en_EP	"The date on which the patent application was first published. It is the date on which the patent document is made available to the public, thereby becoming part of the state of the art." These dates range from 1999-2011.
Year	Espacenet http://worldwide.espacenet.com/?locale=en_EP	This is the year listed in the **priority date** on the citing document.
Immigration	US Department of Homeland	Number of immigrants admitted by country of

	Security http://www.dhs.gov/files/statistics/publications/yearbook.shtm and http://www.dhs.gov/files/statistics/publications/archive.shtm#1	birth, fiscal years 1990-2010.
Lagged Immigration	US Department of Homeland Security http://www.dhs.gov/files/statistics/publications/yearbook.shtm and http://www.dhs.gov/files/statistics/publications/archive.shtm#1	Sum of total **immigration** from the five years prior to **priority date**.
GDP	World Bank http://data.worldbank.org/indicator/NY.GDP.MKTP.CD	"GDP at purchaser's prices is the sum of gross value added by all resident producers in the economy plus any product taxes and minus any subsidies not included in the value of the products. It is calculated without making deductions for depreciation of fabricated assets or for depletion and degradation of natural resources. Data are in current U.S. dollars. Dollar figures for GDP are converted from domestic currencies using single year official exchange rates. For a few countries where the official exchange rate does not reflect the rate effectively applied to actual foreign exchange transactions, an alternative conversion factor is used." Note: GDP for Australia in 2010 was missing and, but found at the IMF website: http://elibrary-data.imf.org/DataReport.aspx?c=1449311&d=33060&e=161838
Trade	U.S. Department of Commerce, Bureau of the Census, Foreign Trade http://www.census.gov/foreign-trade/balance/	Imports of US goods from sample countries plus exports to US in millions of dollars. Values are not seasonally adjusted.
Patent Stock	Espacenet http://worldwide.espacenet.com/?locale=en_EP	Sum of total patents and patent applications, all classes, in sample countries for each year from 1995-2010.
Education 1	UNESCO Institute for Statistics	Enrolment in total tertiary. Public and private. Full and part time. Total.

	http://www.uis.unesco.org	
Education 2	UNESCO Institute for Statistics http://www.uis.unesco.org	School life expectancy (years). Tertiary. Total
Education 3	UNESCO Institute for Statistics http://www.uis.unesco.org	School life expectancy (years). Primary to secondary. Total
FDI	UN Conference on Trade and Development http://unctadstat.unctad.org/TableViewer/tableView.aspx	Inward foreign direct investment stock. US Dollars at current prices and current exchange rates in millions.
Patent Stock Agriculture	Espacenet http://worldwide.espacenet.com/?locale=en_EP	Sum of total patents and patent applications from class A01 of the international patent classification (IPC) scheme. This class includes agriculture; forestry; animal husbandry; hunting; trapping; fishing.
HDI	UN Development Program – Human Development Reports 1998, 2006, 2011	Human Development Index, summary measure of human development. More details, including compilation of index, in body of paper.
% English Speaking	Australia - 2001 Australian Census Belgium, Denmark, Finland, France, Germany, Greece, Netherlands, Spain, Turkey, United Kingdom - Eurobarometer report 2006 Japan – jref.com Switzerland - Federal Statistical Office, Neuchâtel 2008	Percentage of population in each sample **country** that speak English as a first or second language.
Number of English Speakers	Australia - 2001 Australian Census Belgium, Denmark, Finland, France, Germany, Greece, Netherlands, Spain, Turkey, United Kingdom - Eurobarometer report 2006 Japan – jref.com Switzerland - Federal Statistical Office, Neuchâtel 2008	Total number of native English speakers in each sample **country**.
Distance	Geobytes http://www.geobytes.com/citydistancetool.htm	Distance in miles from the capitol of the sample **country** to Washington, D.C.

Appendix D

Results Tables

Table D1. FE OLS, Original Variables

Variable	Coefficient (Standard Error)	t-statistic	95% Confidence Interval	
GDP	6.46e-16 (3.68e-16)	1.75	-7.63e-17	1.37e-15
Trade	5.19e-08 (1.03e-08)	5.05**	3.17e-08	7.21e-08
Patent Stock	-2.17e-08 (3.45e-09)	-6.30**	-2.85e-08	-1.49e-08
Education	.0005314 (.0000717)	7.41**	.0003907	.000672
Immigration	1.38e-08 (1.21e-08)	1.14	-1.00e-08	3.76e-08
Note: **Significant at the 5 percent level. $R^2 = 0.0015$				

Table D2. FE Poisson, Original Variables

Variable	Coefficient (Standard Error)	z-statistic	95% Confidence Interval	
GDP	1.99e-13 (9.11e-14)	2.19**	2.08e-14	3.78e-13
Trade	.0000139 (2.54e-06)	5.45**	8.88e-06	.0000189
Patent Stock	-5.23e-06 (6.49e-07)	-8.05**	-6.50e-06	-3.96e-06
Education	.27509 (.0303498)	9.06**	.2156054	.3345746
Immigration	6.97e-07 (2.19e-06)	0.32	-3.59e-06	4.99e-06
Note: **Significant at the 5 percent level.				

Table D3. Fixed Effects OLS, New Variables, HDI Index

Variable	Coefficient (Standard Error)	t-statistic	95% Confidence Interval	
GDP	-2.71e-16 (3.14e-16)	-0.86	- 8.87e-16	3.46e-16
FDI	4.74e-09 (6.97e-10)	6.81**	3.38e-09	6.11e-09
Trade	1.08e-08 (9.38e-09)	1.16	-7.56e-09	2.92e-08
Patent Stock Ag	2.99e-07 (1.53e-07)	1.95	-1.39e-09	5.99e-07
HDI Index 1	.0024989 (.0011694)	2.14**	.0002047	.004793
HDI Index 2	.0061635 (.0011092)	5.56**	.0039874	.0083396
HDI Index 3	.0037433 (.0010783)	3.47**	.0016279	.0058587
Immigration	6.29e-08 (2.41e-08)	2.61**	1.57e-08	1.10e-07
Constant	-.0038226 (.0009561)	-4.00**	-.0056983	-.0019469

Note: **Significant at the 5 percent level.
$R^2 = 0.0018$

Table D4. Fixed Effects OLS, New Variables, No HDI Index

Variable	Coefficient (Standard Error)	t-statistic	95% Confidence Interval	
GDP	-7.64e-16 (3.16e-16)	-2.42**	-1.38e-15	-1.44e-16
FDI	3.78e-09 (6.48e-10)	5.83**	2.51e-09	5.05e-09
Trade	1.48e-08 (9.36e-09)	1.59	-3.52e-09	3.32e-08
Patent Stock Ag	4.17e-07 (1.52e-07)	2.74**	1.19e-07	7.16e-07
HDI	.0127712 (.0013336)	9.58**	.010155	.0153874
Immigration	8.55e-08 (2.32e-08)	3.68**	3.99e-08	1.31e-07
Constant	-.010392 (.0012036)	-8.63**	-.0127533	-.0080308
Note: **Significant at the 5 percent level. $R^2 = 0.0013$				

Table D5. Fixed Effects Poisson, New Variables, HDI Index

Variable	Coefficient (Standard Error)	z-statistic	95% Confidence Interval	
GDP	-4.65e-14 (1.03e-13)	-0.45	-2.48e-13	1.55e-13
FDI	1.06e-06 (1.63e-07)	6.50**	7.41e-07	1.38e-06
Trade	3.57e-06 (2.60e-06)	1.37	-1.53e-06	8.67e-06
Patent Stock Ag	.0000603 (.0000397)	1.52	-.0000175	.000138
HDI Index 1	3.96721 (.8027314)	4.94**	2.393885	5.540534
HDI Index 2	5.000825 (.7436038)	6.73**	3.543388	6.458261
HDI Index 3	3.885115 (.7870518)	4.94**	2.342522	5.427709
Immigration	.0000106 (4.29e-06)	2.46**	2.14e-06	.000019
Note: **Significant at the 5 percent level.				

Table D6. Fixed Effects Poisson, New Variables, No HDI Index

Variable	Coefficient (Standard Error)	z-statistic	95% Confidence Interval	
GDP	-1.84e-13 (8.87e-14)	-2.07**	-3.58e-13	-1.02e-14
FDI	1.04e-06 (1.58e-07)	6.61**	7.34e-07	1.35e-06
Trade	5.33e-06 (2.36e-06)	2.26**	7.10e-07	9.94e-06
Patent Stock Ag	.0000839 (.0000374)	2.24**	.0000105	.0001573
HDI	10.11297 (1.051712)	9.62**	8.051653	12.17429
Immigration	.0000139 (3.96e-06)	3.50**	6.12e-06	.0000217
Note: **Significant at the 5 percent level.				

Sensitivity Analysis - Immigration Variables[14]

Table D7.1. Original Immigration, Immigration = Total Lags 1-5

Variable	**Coefficient (Standard Error)**	**z-statistic**	**95% Confidence Interval**	
GDP	-2.65e-14 (1.07e-13)	-0.25	-2.36e-13	1.83e-13
FDI	1.20e-06 (1.61e-07)	7.43**	8.80e-07	1.51e-06
Trade	4.98e-07 (2.80e-06)	0.18	-4.98e-06	5.98e-06
Patent Stock Ag	.0001143 (.0000399)	2.86**	.000036	.0001926
HDI Index 1	4.734698 (.9323029)	5.08**	2.907418	6.561978
HDI Index 2	5.648935 (.860779)	6.56**	3.961839	7.336031
HDI Index 3	4.413049 (.8955474)	4.93**	2.657808	6.168289
Immigration	.0000114 (2.54e-06)	4.48**	6.41e-06	.0000164
Note: **Significant at the 5 percent level.				

[14] All tables in this section report results from a FE Poisson Regression Model, New Variables, HDI Index

Table D7.2. Original Immigration, Immigration = Lag 1

Variable	Coefficient (Standard Error)	z-statistic	95% Confidence Interval	
GDP	-1.03e-14 (1.07e-13)	-0.10	-2.21e-13	2.00e-13
FDI	1.12e-06 (1.67e-07)	6.74**	7.97e-07	1.45e-06
Trade	9.57e-07 (2.75e-06)	0.35	-4.43e-06	6.35e-06
Patent Stock Ag	.0000978 (.0000389)	2.52**	.0000217	.000174
HDI Index 1	4.593508 (.9088854)	5.05**	2.812125	6.374891
HDI Index 2	5.478111 (.8373602)	6.54**	3.836915	7.119307
HDI Index 3	4.295844 (.8758718)	4.90**	2.579167	6.012521
Immigration Lag 1	.0000504 (.0000115)	4.38**	.0000279	.000073

Note: **Significant at the 5 percent level.

Table D7.3. Original Immigration, Immigration = Lag 2

Variable	Coefficient (Standard Error)	z-statistic	95% Confidence Interval	
GDP	-5.10e-14 (1.06e-13)	-0.48	-2.59e-13	1.57e-13
FDI	1.24e-06 (1.63e-07)	7.59**	9.19e-07	1.56e-06
Trade	2.89e-06 (2.66e-06)	1.09	-2.32e-06	8.10e-06
Patent Stock Ag	.0000892 (.0000403)	2.21**	.0000101	.0001682
HDI Index 1	3.995195 (.9102468)	4.39**	2.211144	5.779246
HDI Index 2	4.962569 (.8377745)	5.92**	3.320561	6.604577
HDI Index 3	3.78566 (.8728603)	4.34**	2.074885	5.496435
Immigration Lag 2	.0000366 (.0000117)	3.13**	.0000137	.0000596
Note: **Significant at the 5 percent level.				

Table D7.4. Original Immigration, Immigration = Lag 3

Variable	Coefficient (Standard Error)	z-statistic	95% Confidence Interval	
GDP	-7.78e-14 (1.07e-13)	-0.73	-2.88e-13	1.32e-13
FDI	1.30e-06 (1.59e-07)	8.18**	9.91e-07	1.61e-06
Trade	1.97e-06 (2.63e-06)	0.75	-3.18e-06	7.11e-06
Patent Stock Ag	.0001146 (.0000401)	2.85**	.0000359	.0001933
HDI Index 1	4.320163 (.9237784)	4.68**	2.50959	6.130735
HDI Index 2	5.306819 (.8542184)	6.21**	3.632582	6.981057
HDI Index 3	4.049474 (.8841363)	4.58**	2.316598	5.782349
Immigration Lag 3	. 0000508 (.000011)	4.62**	.0000292	.0000723
Note: **Significant at the 5 percent level.				

Table D7.5. Original Immigration, Immigration = Lag 4

Variable	Coefficient (Standard Error)	z-statistic	95% Confidence Interval	
GDP	-2.26e-14 (1.06e-13)	-0.21	-2.30e-13	1.85e-13
FDI	1.28e-06 (1.57e-07)	8.14**	9.69e-07	1.58e-06
Trade	5.51e-07 (2.69e-06)	0.21	-4.72e-06	5.82e-06
Patent Stock Ag	. 0001164 (.0000393)	2.96**	.0000394	.0001934
HDI Index 1	4.553044 (.9151606)	4.98**	2.759362	6.346726
HDI Index 2	5.538291 (.8469695)	6.54**	3.878262	7.198321
HDI Index 3	4.289058 (.8865779)	4.84**	2.551398	6.026719
Immigration Lag 4	.0000535 (9.71e-06)	5.51**	.0000345	.0000725
Note: **Significant at the 5 percent level.				

Table D7.6. Original Immigration, Immigration = Lag 5

Variable	Coefficient (Standard Error)	z-statistic	95% Confidence Interval	
GDP	-1.84e-14 (1.04e-13)	-0.18	-2.23e-13	1.86e-13
FDI	1.28e-06 (1.59e-07)	8.03**	9.67e-07	1.59e-06
Trade	2.64e-06 (2.74e-06)	0.96	-2.74e-06	8.01e-06
Patent Stock Ag	.0000837 (.0000406)	2.06**	4.17e-06	.0001633
HDI Index 1	3.899891 (.8633575)	4.52**	2.207741	5.592041
HDI Index 2	4.888436 (.7980454)	6.13**	3.324296	6.452576
HDI Index 3	3.741401 (.8401556)	4.45**	2.094726	5.388076
Immigration Lag 5	.0000301 (.0000105)	2.88**	9.59e-06	.0000506
Note: **Significant at the 5 percent level.				

Table D7.7. New Immigration, Immigration = Total Lags 1-5

Variable	Coefficient (Standard Error)	z-statistic	95% Confidence Interval	
GDP	-4.65e-14 (1.03e-13)	-0.45	-2.48e-13	1.55e-13
FDI	1.06e-06 (1.63e-07)	6.50**	7.41e-07	1.38e-06
Trade	3.57e-06 (2.60e-06)	1.37	-1.53e-06	8.67e-06
Patent Stock Ag	.0000603 (.0000397)	1.52	-.0000175	.000138
HDI Index 1	3.96721 (.8027314)	4.94**	2.393885	5.540534
HDI Index 2	5.000825 (.7436038)	6.73**	3.543388	6.458261
HDI Index 3	3.885115 (.7870518)	4.94**	2.342522	5.427709
Immigration	.0000106 (4.29e-06)	2.46**	2.14e-06	.000019
Note: **Significant at the 5 percent level.				

Table D7.8. New Immigration, Immigration = Lag 1

Variable	Coefficient (Standard Error)	z-statistic	95% Confidence Interval	
GDP	-2.43e-14 (1.04e-13)	-0.23	-2.28e-13	1.79e-13
FDI	1.20e-06 (1.66e-07)	7.23**	8.73e-07	1.52e-06
Trade	4.20e-06 (2.56e-06)	1.64	-8.24e-07	9.23e-06
Patent Stock Ag	.000058 (.000039)	1.49	-.0000185	.0001345
HDI Index 1	3.398244 (.8334641)	4.08**	1.764684	5.031804
HDI Index 2	4.429488 (.769695)	5.75**	2.920914	5.938062
HDI Index 3	3.353884 (.8152195)	4.11**	1.756083	4.951685
Immigration Lag 1	.0000435 (.0000194)	2.24**	5.46e-06	.0000815
Note: **Significant at the 5 percent level.				

Table D7. 9. New Immigration, Immigration = Lag 2

Variable	Coefficient (Standard Error)	z-statistic	95% Confidence Interval	
GDP	-5.27e-14 (1.01e-13)	-0.52	-2.51e-13	1.46e-13
FDI	1.36e-06 (1.65e-07)	8.24**	1.03e-06	1.68e-06
Trade	6.18e-06 (2.49e-06)	2.48**	1.30e-06	.0000111
Patent Stock Ag	.0000439 (.0000404)	1.09	-.0000353	.0001231
HDI Index 1	2.766798 (.8339697)	3.32**	1.132247	4.401348
HDI Index 2	3.887428 (.7691272)	5.05**	2.379966	5.39489
HDI Index 3	3.353884 (2.818463)	3.49**	1.234782	4.402143
Immigration Lag 2	-5.56e-07 (.8080153)	-0.03	-.000043	.0000419
Note: **Significant at the 5 percent level.				

Table D7.10. New Immigration, Immigration = Lag 3

Variable	Coefficient (Standard Error)	z-statistic	95% Confidence Interval	
GDP	-4.58e-14 (1.04e-13)	-0.44	-2.50e-13	1.59e-13
FDI	1.23e-06 (1.61e-07)	7.63**	9.11e-07	1.54e-06
Trade	4.13e-06 (2.50e-06)	1.65	-7.73e-07	9.04e-06
Patent Stock Ag	.0000659 (.0000396)	1.67	-.0000116	.0001434
HDI Index 1	3.490932 (.8445433)	4.13**	1.835657	5.146206
HDI Index 2	4.568679 (.7834528)	5.83**	3.033139	6.104218
HDI Index 3	3.370286 (.8200953)	4.11**	1.762929	4.977644
Immigration Lag 3	.0000561 (.0000217)	2.58**	.0000135	.0000987
Note: **Significant at the 5 percent level.				

Table D7.11. New Immigration, Immigration = Lag 4

Variable	Coefficient (Standard Error)	z-statistic	95% Confidence Interval	
GDP	-1.82e-14 (1.04e-13)	-0.18	-2.21e-13	1.85e-13
FDI	1.23e-06 (1.58e-07)	7.79**	9.18e-07	1.53e-06
Trade	3.43e-06 (2.53e-06)	1.36	-1.52e-06	8.39e-06
Patent Stock Ag	.000068 (.0000387)	1.76	-7.85e-06	.0001439
HDI Index 1	3.64943 (.8532019)	4.28**	1.977185	5.321675
HDI Index 2	4.727529 (.7903725)	5.98**	3.178428	6.276631
HDI Index 3	3.630424 (.8357493)	4.34**	1.992385	5.268462
Immigration Lag 4	.0000635 (.0000189)	3.36**	.0000265	.0001005
Note: **Significant at the 5 percent level.				

Table D7.12. New Immigration, Immigration = Lag 5

Variable	Coefficient (Standard Error)	z-statistic	95% Confidence Interval	
GDP	-7.06e-14 (1.01e-13)	-0.70	-2.68e-13	1.27e-13
FDI	1.41e-06 (1.63e-07)	8.64**	1.09e-06	1.73e-06
Trade	7.19e-06 (2.54e-06)	2.83**	2.21e-06	.0000122
Patent Stock Ag	.0000356 (.0000404)	0.88	-.0000435	.0001147
HDI Index 1	2.640842 (.7912022)	3.34**	1.090114	4.19157
HDI Index 2	3.78408 (.7316788)	5.17**	2.350016	5.218144
HDI Index 3	2.779566 (.7787512)	3.57**	1.253242	4.30589
Immigration Lag 5	-.0000225 (.0000217)	-1.04	-.000065	.0000201
Note: **Significant at the 5 percent level.				

Sensitivity Analysis - Regressors[15]

Table D8.1. Regressors: None

Variable	Coefficient (Standard Error)	z-statistic	95% Confidence Interval	
Immigration	.0000316 (2.23e-06)	14.15**	.0000272	.000036
Note: **Significant at the 5 percent level.				

Table D8.2. Regressors: GDP, FDI, Trade

Variable	Coefficient (Standard Error)	z-statistic	95% Confidence Interval	
GDP	-2.25e-13 (8.11e-14)	-2.77**	-3.84e-13	-6.57e-14
FDI	9.44e-07 (1.20e-07)	7.85**	7.08e-07	1.18e-06
Trade	.0000111 (2.07e-06)	5.35**	7.00e-06	.0000151
Immigration	5.92e-06 (3.73e-06)	1.59	-1.39e-06	.0000132
Note: **Significant at the 5 percent level.				

[15] All tables in this section report results from a FE Poisson Regression Model, New Variables, HDI Index.

Table D8.3. Regressors: GDP, FDI, Patent Stock

Variable	Coefficient (Standard Error)	z-statistic	95% Confidence Interval	
GDP	-1.14e-13 (5.49e-14)	-2.08**	-2.22e-13	-6.76e-15
FDI	1.38e-06 (1.54e-07)	8.97**	1.08e-06	1.68e-06
Patent Stock Ag	.0001898 (.0000312)	6.09**	.0001287	.0002509
Immigration	.0000153 (3.64e-06)	4.21**	8.20e-06	.0000225
Note: **Significant at the 5 percent level.				

Table D8.4. Regressors: GDP, FDI, HDI

Variable	Coefficient (Standard Error)	z-statistic	95% Confidence Interval	
GDP	1.93e-13 (2.69e-14)	7.16**	1.40-13	2.45e-13
FDI	7.24e-07 (1.16e-07)	6.23**	4.96e-07	9.52e-07
HDI Index 1	4.80613 (.8642619)	5.56**	3.112208	6.500052
HDI Index 2	5.835158 (.7989143)	7.30**	4.269315	7.401002
HDI Index 3	4.622473 (.8322774)	5.55**	2.991239	6.253706
Immigration	.0000117 (3.75e-06)	3.13**	4.38e-06	.0000191
Note: **Significant at the 5 percent level.				

Table D8.5. Regressors: GDP, Trade, Patent Stock

Variable	Coefficient (Standard Error)	z-statistic	95% Confidence Interval	
GDP	-7.83e-14 (8.87e-14)	-0.88	-2.52e-13	9.56e-14
Trade	7.82e-06 (2.55e-06)	3.07**	2.83e-06	.0000128
Patent Stock Ag	-.0000303 (.0000323)	-0.94	-.0000936	.000033
Immigration	.0000197 (4.06e-06)	4.86**	.0000118	.0000277
Note: **Significant at the 5 percent level.				

Table D8.6. Regressors: GDP, Trade, HDI

Variable	Coefficient (Standard Error)	z-statistic	95% Confidence Interval	
GDP	2.27e-13 (9.69e-14)	2.34**	3.73e-14	4.17e-13
Trade	-1.77e-06 (2.32e-06)	-0.76	-6.32e-06	2.78e-06
HDI Index 1	7.220177 (.9762179)	7.40**	5.306825	9.133529
HDI Index 2	7.985929 (.9070067)	8.80**	6.208229	9.76363
HDI Index 3	6.594753 (.9344446)	7.06**	4.763276	8.426231
Immigration	.0000237 (3.01e-06)	7.87**	.0000178	.0000296
Note: **Significant at the 5 percent level.				

Table D8.7. Regressors: GDP, Patent Stock, HDI

Variable	Coefficient (Standard Error)	z-statistic	95% Confidence Interval	
GDP	2.63e-13 (6.01e-14)	4.37**	1.45e-13	3.81e-13
Patent Stock Ag	-.0000572 (.0000301)	-1.90	-.0001161	1.67e-06
HDI Index 1	6.91356 (.9788733)	7.06**	4.995003	8.832116
HDI Index 2	7.760683 (.9033238)	8.59**	5.990201	9.531165
HDI Index 3	6.399895 (.9303162)	6.88**	4.576509	8.223282
Immigration	.0000189 (3.95e-06)	4.79**	.0000112	.0000266
Note: **Significant at the 5 percent level.				

Table D8.8. Regressors: FDI, Trade, Patent Stock

Variable	Coefficient (Standard Error)	z-statistic	95% Confidence Interval	
FDI	1.15e-06 (1.48e-07)	7.80**	8.62e-07	1.44e-06
Trade	1.13e-06 (1.63e-06)	0.69	-2.07e-06	4.33e-06
Patent Stock Ag	.0001127 (.0000365)	3.09**	.0000413	.0001842
Immigration	.0000117 (4.05e-06)	2.89**	3.77e-06	.0000196
Note: **Significant at the 5 percent level.				

Table D8.9. Regressors: FDI, Trade, HDI

Variable	Coefficient (Standard Error)	z-statistic	95% Confidence Interval	
FDI	8.79e-07 (1.21e-07)	7.26**	6.41e-07	1.12e-06
Trade	4.96e-06 (6.60e-07)	7.51**	3.66e-06	6.25e-06
HDI Index 1	4.226043 (.8593229)	4.92**	2.541801	5.910285
HDI Index 2	5.273448 (.7949828)	6.63**	3.715311	6.831586
HDI Index 3	4.148149 (.8268969)	5.02**	2.52746	5.768837
Immigration	8.60e-06 (3.95e-06)	2.18**	8.64e-07	.0000163
Note: **Significant at the 5 percent level.				

Table D8.10. Regressors: FDI, Patent Stock, HDI

Variable	Coefficient (Standard Error)	z-statistic	95% Confidence Interval	
FDI	1.11e-06 (1.31e-07)	8.48**	8.55e-07	1.37e-06
Patent Stock Ag	.0001067 (.0000145)	7.37**	.0000784	.0001351
HDI Index 1	4.185859 (.8228177)	5.09**	2.573166	5.798552
HDI Index 2	5.185144 (.7652378)	6.78**	3.685305	6.684982
HDI Index 3	4.033946 (.8022743)	5.03**	2.461517	5.606375
Immigration	.0000145 (3.52e-06)	4.11**	7.57e-06	.0000214
Note: **Significant at the 5 percent level.				

Table D8.11. Regressors: Trade, Patent Stock, HDI

Variable	Coefficient (Standard Error)	z-statistic	95% Confidence Interval	
Trade	5.79e-06 (1.67e-06)	3.47**	2.52e-06	9.06e-06
Patent Stock Ag	-.0000532 (.0000348)	-1.53	-.0001214	.000015
HDI Index 1	6.990433 (.9838425)	7.11**	5.062137	8.918729
HDI Index 2	7.755863 (.9120627)	8.50**	5.968253	9.543473
HDI Index 3	6.457579 (.9365344)	6.90**	4.622005	8.293153
Immigration	.0000193 (4.28e-06)	4.51**	.0000109	.0000277
Note: **Significant at the 5 percent level.				

New Panel Model[16]

Table D9.1. FE OLS, New Variables, HDI Index

Variable	Coefficient (Standard Error)	t-statistic	95% Confidence Interval	
GDP	-4.03e-12 (2.48e-12)	-1.63	-9.43e-12	1.36e-12
FDI	5.20e-06 (3.71e-06)	1.40	-2.87e-06	.0000133
Trade	-.0000196 (.000015)	-1.31	-.0000524	.0000131
Patent Stock Ag	.0021771 (.0003155)	6.90**	.0014896	.0028646
HDI Index 1	29.54556 (15.55089)	1.90	-4.336913	63.42803
HDI Index 2	29.21662 (15.00282)	1.95	-3.471716	61.90495
HDI Index 3	24.90603 (14.70241)	1.69	-7.127763	56.93983
Immigration	-.0000424 (.0001718)	-0.25	-.0004166	.0003319
Constant	-19.87755 (14.21884)	-1.40	-50.85774	11.10264
Note: **Significant at the 5 percent level. $R^2 = 0.3210$				

[16] All tables in this section report results from a FE Poisson Regression Model, New Variables, HDI Index

Table D9.2. FE Poisson, New Variables, HDI Index

Variable	Coefficient (Standard Error)	z-statistic	95% Confidence Interval	
GDP	-6.46e-13 (2.45e-13)	-2.64**	-1.13e-12	-1.66e-13
FDI	5.97e-07 (5.41e-07)	1.10	-4.62e-07	1.66e-06
Trade	-3.17e-06 (5.77e-06)	-0.55	-.0000145	8.14e-06
Patent Stock Ag	.0006656 (.0000627)	10.61**	.0005426	.0007885
HDI Index 1	13.09801 (5.937998)	2.21**	1.459752	24.73628
HDI Index 2	12.72772 (5.605145)	2.27**	1.741836	23.7136
HDI Index 3	11.27374 (5.547188)	2.03**	.4014531	22.14603
Immigration	-2.24e-07 (.0000178)	-0.01	-.000035	.0000346
Note: **Significant at the 5 percent level.				

References

Agrawal, Ajay, Iain Cockburn and John McHale. "Gone but Not Forgotten: Labor Flows, Knowledge Spillovers and Enduring Social Capital". NBER Working Paper, No. 9950: pp.1-35. September 2003.

Audretsch, David B. "Agglomeration and the Location of Innovative Activity". *Oxford Review of Economic Policy*, 14(2): pp.18-29. 1998.

Bodvarsson, Orn B. and Hendrik Van den Berg. *The Economics of Immigration: Theory and Policy*. Springer-Verlag Berlin Heidelberg. 2009.

Cooley, Thomas F. and LeRoy, Stephen F. "Identification and Estimation *of* Money Demand." *American Economic Review*, 71: pp.825-844. December 1981.

Diplomatic Conference for the Adoption of an Agreement on the Creation of an Industrial
Property Organization for English-Speaking Africa at Lusaka (Zambia). "Lusaka Agreement on the Creation of the African Regional Intellectual Property Organization (ARIPO)." December 9, 1976. Treaties and Other International Agreements.

Dolado, Juan, Alessandra Goria and Andrea Ichino. "Immigration, Human Capital and Growth in the Host Country: Evidence from Pooled Country Data". *Journal of Population Economics*, 7(2): pp.193-215. June 1994.

Florida, Richard. "The World is Spiky". *The Atlantic Monthly*. October 2005.

Galor, Oded and Oded Stark. "The Impact of Differences in the Levels of Technology on International Labor Migration". *Journal of Population Economics*, 4(1): pp.1-12. March 1991.

Haque, Nadeem U. and Se-Jik Kim. "'Human Capital Flight': Impact of Migration on Income and Growth". *Staff Papers – International Monetary Fund*, 42(3): pp.577-607. September 1995.

Hausman, Jerry, Bronwyn H. Hall and Zvi Griliches. "Econometric Models for Count Data with an Application to the Patents-R & D Relationship." *Econometrica*, 53(4): pp.909-938. July 1984.

Huber, P. J. "The behavior of maximum likelihood estimates under nonstandard conditions." *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, Vol 1: pp.221–233. Berkeley: University of California Press. 1967.

"International Migration Outlook – SOPEMI 2008." OECD. September 2008.

Jaffe, Adam B., Manuel Trajtenberg and Rebecca Hendersen. "Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations". *The Quarterly Journal of Economics*, 108(3): pp.577-598. August 1993.

Jolly, Richard et al. *Human Development Report 1998: Consumption for Human Development*.United Nations Development Program. Oxford University Press. New York, NY. 1998.

Kondo, Hitoshi. "International Factor Mobility and Production Technology". *Journal of Population Economics*, 2(4): pp.281-299. December 1989.

Klugman, Jeni et al. *Human Development Report 2011: Sustainability and Equity: A Better Future for All*. United Nations Development Program. Oxford University Press. NewYork, NY. 2011.

Krugman, Paul. *Geography and Trade*. The MIT Press. Cambridge, MA. 1991.

Leamer, E. E. and H. Leonard. "Reporting the Fragility of Regression Estimates." *Review of Economics and Statistics*, 65: pp.306-317. 1983.

Levine, Ross and David Renelt. "A Sensitivity Analysis of Cross-Country Growth Regression". *American Economic Association*, 82(4): pp.942-963. September 1992.

Mayr, Karin and Giovanni Peri. "Return Migration as a Channel of 'Brain Gain'". NBER Working Paper No. 14039: pp.1-52. May 2008.

Miyagiwa, Kaz. "Scale Economies in Education and the 'brain drain' Problem". *International Economic Review*, 32(3): pp.743-759. August 1991.

Murakami, Yukiko. "Japan's Brain Drain: An Analysis of Japanese Researchers Living in the United States." *The Japanese Economy*, 37(2): pp.23-57. Summer 2010.

Ottaviano, Gianmarco I.P. and Diego Puga. "Agglomeration in the Global Economy: A Survey in the 'New Economic Geography'". Programme on International Economic Performance. pp.707-731. 1998.

Saxenian, AnnaLee. "Brain Circulation: How High-Skill Immigration Makes Everyone Better Off". *The Brookings Review*, 20(1): pp.28-31. Winter 2002.

Sjöholm, Fredrik. "International Transfer of Knowledge: The Role of International Trade and Geographic Proximity." *Weltwirtschaftliches*, 132: pp.97-115. 1996.

Saint-Paul, Gilles. “The Brain Drain: Some Evidence from European Expatriates in the United States.” IZA Discussion Paper No. 1310. September 2004.

StataCorp. “Stata: Release 11.” *Statistical Software.* College Station, TX: StataCorp LP. 2009.

US Immigration and Naturalization Service. *1997 Statistical Yearbook of the Immigration and Naturalization Service*. US Government Printing Office. Washington, D.C. 2009.

Van den Berg, Hendrik and Joshua J. Lewer. *International Trade and Economic Growth*.M.E. Sharpe. Armonk, NY. 2007.

Washington Diplomatic Conference on the Patent Cooperation Treaty. “Patent Cooperation Treaty.” June 19, 1970. Treaties and Other International Agreements.

Watkins, Kevin et al. *Human Development Report 2006: Beyond Scarcity: Power, Poverty, and the Global Water Crisis*.United Nations Development Program. Oxford University Press. New York, NY. 2006.

White, H. “A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity.” *Econometrica,* 48: pp.817–838. 1980.

White, H. “Maximum Likelihood Estimation of Misspecified Models.” *Econometrica,* 50: pp.1–25.1982.

Wooldridge, Jeffrey M. *Econometric Analysis of Cross Section and Panel Data*. The MIT Press. Cambridge, MA. 2002.