Doing R&D in Countries with Weak IPR Protection: Can Corporate Management Substitute for Legal Institutions?*

Minyuan Zhao
Stern School of Business, NYU
44 West 4th Street, KMC 7-176
New York, NY 10012
Tel: (212) 998-0881
mzhao@stern.nyu.edu

April 8, 2004

* This paper is developed under the supervision of Professor Bernard Yeung. I am especially grateful to Juan Alcacer, Rachelle Sampson and William Greene for their continuous support. I also thank Bharat Anand, Raymond Fisman, Witold Henisz, Tarun Khanna, Steve Klepper, Joanne Oxley, Jordan Siegel, Olav Sorenson, Mary Tripsas, Sidney Winter, Rosemarie Ziedonis and seminar participants at Carnegie Mellon University, New York University, CCC, HBS Strategy Conference and Organization Science Fall Conference for valuable comments and suggestions. All errors remain my own. I appreciate the financial support from Willensky Fellowship, AIB-CIBER dissertation proposal grant, and Stern School of Business dissertation research fund.

Doing R&D in Countries with Weak IPR Protection: Can Corporate Management Substitute for Legal Institutions?

ABSTRACT

Multinational enterprises (MNEs) are increasingly conducting R&D in countries such as India and China, where intellectual property rights (IPR) protection is still weak. This paper examines the puzzle. The argument is that weak IPR leads to low returns to innovation and thus underutilization of innovative talents. MNEs who possess not only the capabilities to utilize these talents, but also the internal organizations to protect the intellectual properties will therefore find it attractive to conduct R&D at those locations. Following a series of interviews in major multinational R&D centers in China, a theoretical framework is presented to capture the interaction between firm strategies and institutional environment. Empirical findings from a sample of 1,567 U.S.-headquartered innovating firms are consistent with the hypotheses that (i) technologies developed in weak IPR countries are used more internally, and (ii) firms doing R&D in weak IPR countries have tighter internal technology structures. The results suggest that firms may be using strong internal linkages to substitute for the inadequate external institutions. By doing so, they are able to take advantage of the arbitrage opportunities presented by the institutional gap across countries.

Keywords: R&D, intellectual property rights, MNE, arbitrage

> Intellectual property is still an extremely vague concept in China, where fake DVDs are sold on street corners and even the Government uses pirated software.
>
> –– The Times (London), Dec. 12, 2002

> A significant number of multinationals are increasingly combing the mainland [China] for engineers and researchers to handle projects for global applications that, in recent years, would have been performed in labs in the United States or Europe.
>
> –– ZDNet News, Jul. 10, 2002

I. INTRODUCTION

The recent years have witnessed a surge of multinational R&D activity in countries such as India and China, where the intellectual property rights (IPR) protection is still far from satisfactory. Technology giants Microsoft, IBM, Intel, and GE are in the lead, but more firms are following. (Financial Times 4/19/02; New York Times 4/21/02; BusinessWeek 2/03/03) Moreover, the R&D conducted in these Indian and Chinese labs is in excess of that required for product localization or government-enforced technology transfers.

This trend is in apparent contradiction of conventional wisdom. Because poor institutional environment erodes the appropriable value of innovation, firms have been advised to keep their knowledge-intensive activities away from weak IPR countries. What has enabled some firms to act differently?

To understand this puzzle, I began with a series of interviews with the researchers and managers in some of the multinational R&D labs in China. Some common practices emerged, including intensive communications and collaborations with headquarters, patent applications in the home country, and internal project transfers across countries. In particular, the projects are often closely integrated in the firms' global research agendas. The "carved-out expertise" – as called by some labs – is valuable only when combined with the complementary knowledge and resources within the firm. Even if imitation occurs, the value that can be taken away from the firm is very limited. The closely-knit internal innovation structure, therefore, serves as an immune system against the adverse external environment.

This observation suggests a framework to examine the original puzzle. In countries with poor IPR protection and poor institutional environment overall, local firms would find it difficult to appropriate value from intellectual products. As a result, R&D is discouraged and human capital is undervalued. This is in spite of the fact that these countries have a large pool of potentially valuable talent to conduct R&D. MNEs are in a unique position to arbitrage the difference in factor prices across national borders; their ability to do so stems from their internal organizations that can be viewed as a substitute for the inadequate external institutions. I call this the *internalization-arbitrage* conjecture.

To articulate the interaction between firm organizations and external environment, I map the observations to a very simple, but logically contestable framework. It shows that MNEs may find it desirable to conduct R&D across borders when technologies are complementary internally. By keeping the complementary resources well protected, MNEs are able to leverage the strong institutions in the home country for their operations overseas. The viability of this strategy depends on a set of firm-specific and knowledge-specific characteristics.

I then seek empirical evidence of the theoretical conjecture, using U.S. patent data and the Directory of Corporate Affiliations. I find supportive results that technologies developed in weak IPR countries are used more internally than those developed in other foreign countries. In addition, firms doing R&D in weak IPR countries feature significantly stronger internal linkages than those who do not. The results are consistent with the thought that the internal linkages allow firms to appropriate value from their knowledge even in weak institutional environments.

The following section briefly describes my interviews in China. Section III sets up the theoretical framework and analyzes the potential arbitrage opportunities between different institutional environments. Section IV brings the theoretical conjecture to the data and sets up the stage for empirical analysis. The results are analyzed in Section V. Potential caveats and robustness checks are discussed in Section VI. Section VII concludes and discusses future extensions.

II. THE CHINA STORY

Chinese colleges and universities are churning out nearly half a million science and engineering degrees every year, almost the same number as in the United States. However, domestic R&D investment remains at a very low level; the still weak enforcement of IPR protection is often to blame.

In contrast, MNEs are streaming into China to tap the best talents in the country. By the end of 2002, more than four hundred foreign owned R&D centers have been established, hiring away the best masters and PhDs from China's top universities and research institutes. What enable the MNEs to succeed, besides their deeper pockets and technology advantage, where the locals fail? With this puzzle in mind, I conducted a series of interviews with researchers and managers in some large multinational R&D labs in Beijing and Shanghai in the summer of 2002.[1] The observations suggest a key role of firms' internal organizations.

2.1 Organizational Structure

To utilize and appropriate value from the human capital in China, MNEs have to keep knowledge leakage at the lowest level possible. At the same time, China has a booming domestic market. Competition for market share requires that MNEs effectively transfer and adapt technologies for their local operations. This dilemma leads to a common practice among MNEs in China: the separation of localization-oriented R&D centers and research labs that aim to develop frontier technologies for global applications.

During the interviews, it was repeatedly emphasized to me that the research labs are a *coherent part* of the firms' *worldwide* R&D forces. For example, the goal of Microsoft Research (MSR) Asia in Beijing is to "attract the most talented researchers in the field of computing" and to "advance the state-of-the-art in computer science research." IBM China Research Lab (CRL) has a mission to "create world-class information technologies and the underlying science which propel the world advances." Intel China Software Lab

[1] Information has been updated through emails and phone calls during the past year.

(ICSL) hopes to “create and enhance global value of Intel's silicon, platforms and solutions by delivering innovative software technology and quality products.”

With such missions, a research lab is usually a parallel organization outside of the MNE’s local operations. It reports directly to the technology department at headquarters, connected with the firm’s other labs by intranet, conference calls, regular meetings, and project collaborations. Because the resultant technologies are aimed at global applications, intellectual property issues are mostly handled in the home country.

The centralized organization facilitates the transfer of research projects across locations, and therefore makes it possible to exercise the strategy of differentiated project assignments. Not only do the labs conduct very specific types of R&D, but also R&D projects at specific stages. For example, once a project gets close to commercialization, it may be considered “too risky” to stay in China. The firm will either intensify the monitoring or transfer the project to other locations. In other words, with the tight internal organization of R&D activities, the firm can make sure that the R&D activities in China do not expose too much value to risk.

2.2 The Internal Linkages

Unlike those in the localization-oriented R&D centers, the interviewees in the research labs did not seem to be concerned about imitation risks. When I asked the question: “Given the weak IPR protection in China, are you concerned that the technologies would soon be stolen?” there are mainly two types of responses.

First, “they don’t have the ability to steal.” Researchers believe that the projects in their labs draw heavily on firm-specific expertise, a resource that can only accumulate over time inside the firm. For instance, the “Personalized Cartoon Generation and Animation” project in MSR Asia, which is later used in the most recent version of MSN Messenger (Japan), is a sophisticated technology built on Microsoft’s strength in computer vision and computer graphics. In the Intel lab, most efforts are focused on developing new BIOS, compilers, and device drivers for the Intel architecture platforms. “They are built on Intel technology, and they are part of Intel technology," a team leader said, "Imitation? Not an easy task, especially if the imitators don't have the same exposure."

It is true that copycats do not have to understand the technology before they make copies of the final products. However, few final products are developed in these Beijing or Shanghai labs. The research results will be integrated into the final applications somewhere else, most often at headquarters. What can be taken away from the research labs are abstract algorithms, theoretical development, and experiment reports.

Hence, there came the second type of answers: "why would they steal?" Intel BIOS, of course, can only be used on Intel chips. A major success of MSR Asia in 2002 was "AutoMovie", a technology that can intelligently generate edited movies from home videos. It was later integrated into "Microsoft Movie Maker," which is distributed with the new Windows operating systems. Similar examples are the "Mobile HTML Optimizer" used in Microsoft FrontPage and the "Ink Parsing" technology used in Tablet PC. These are all considered major contributions in the field, but they themselves do not bring direct commercial value to potential imitators. "We don't count on the legal system for protection; we count on the technologies to protect themselves," a researcher said.

In sum, if technologies inside the firm are highly complementary, the leakage of a particular technology should not significantly affect firm value. The two types of answers described above suggest two potential sources of internal complementarity:

- Because the *generation* of these technologies relies heavily on the firms' internal expertise, imitation is difficult without the context.
- Because the *appropriation* of these technologies needs other internal knowledge and resources, the individual technologies do not bring direct value to the imitators.

2.3 The Time Trend

Why is this kind of R&D arrangement such a recent phenomenon? First, China – like other countries with relatively weak IPR regimes – has just opened its doors to foreign investors. Even after the initial opening up, the government required minimum local stakes in foreign invested enterprises, which almost ruled out the possibility of close integration. In fact, nearly all these research labs were established after the removal of government restrictions on wholly owned subsidiaries.

Another important driving force is the development of information technologies (IT), which has dramatically reduced the cost of international coordination. Firms need very intensive communications to make sure that the components developed in China fit seamlessly to the needs at the firm level. These efforts have been greatly facilitated by Internet and the improvement in the local IT infrastructure.

MNEs have also gained more experience in organizing large-scale R&D projects and in dealing with the institutional idiosyncrasy in China. "They are gradually learning how to move smartly, and some MNEs learned their lessons the hard way," a researcher said when referring to the firm's earlier losses from counterfeits.

III. THE THEORETICAL FRAMEWORK

The interviews suggest that firms may be able to use their internal organizations to protect knowledge, hence realizing the potential value of human capital in weak IPR countries. In this section, I map this idea to a logically contestable framework and show how technology complementarities, firm organizations, and legal institutions interact with one another. This framework serves two purposes: to help me study internalization-arbitrage in a structured manner, and to motivate the empirical study in the next section.

3.1 The Nature of Knowledge Diffusion

There are three critical steps in imitation: the motivation to imitate, the ability to imitate, and the possibility of getting around legal restrictions against imitation. In an institutional environment where the legal restrictions barely exist or are not effectively enforced, the first two factors can play a critical role in firms' IPR protection. They both stem from the very nature of knowledge flow.

First, the motivation to imitate is low when technologies are highly dependent on internal resources. Imitation is costly (Mansfield et al. 1981), so it will happen only when imitators can profit from the technologies. Teece (1986) points out that specialized and co-specialized complementary assets are critically important to the successful commercialization of an innovation. Thus, innovators can discourage imitation by developing technologies that require complementary knowledge not readily available to potential imitators. For example, basic research that is still far from commercialization, or technologies that are highly firm specific, are usually less attractive to imitators.

Second, the acquisition of complementary knowledge is subject to the constraints of geographic distance. It has long been realized that a multinational corporation is a geographically distributed innovation network, with the capacity to assimilate, generate and integrate knowledge on a worldwide basis (Bartlett and Ghoshal 1990). Knowledge that is difficult to codify or teach can be more efficiently transferred within the firm (Kogut and Zander 1993). Therefore, outside firms would have to face much higher costs to obtain complementary knowledge across country borders, if not altogether impossible.

From this perspective, the nature of knowledge creation and diffusion presents an opportunity for MNEs to overcome the weak institutional environment in the host country. On one hand, the internal complementarity of technologies makes the leakage of individual components less threatening. On the other hand, the constraints on cross-border knowledge flows enable MNEs to keep the critical knowledge under the protection of home institutions. The combination of these two makes R&D in weak IPR countries a feasible strategy. I will elaborate this idea in the following model.

3.2 Internal Complementarity

To focus on the *organization* of R&D activities, the analysis assumes away other factors that may affect the generation and appropriation of intellectual properties. For instance, large MNEs, with their worldwide production and marketing networks, are in a better position to use the new technologies in a large scale and a broad scope (Cohen and Klepper 1996). They may also enjoy increasing returns to scale in innovation and at the same time face fewer financial constraints. Important as they are, these factors are not essential for this analysis.

Suppose that firm *A* has two technologies under development: a_1 and a_2. Standing alone, in the absence of the other, each a_i ($i = 1, 2$) has value $V(a_i)$. For example, Sharp's CG-Silicon technology provides higher resolution and brightness to compact LCDs. Hence, it has stand-alone value $V(a_1)$, which can be considered the rent that Sharp is able to collect from LCD manufacturers for the use of a_1. Once imitation occurs, Sharp would lose its pricing power and $V(a_1)$ would plummet to zero.

At the same time, a_1 can be integrated with a_2 inside the firm. When Sharp combines the CG-Silicon technology with its leading strength in TFT-LCD, it creates the "smart" displays that make possible a new generation of feature-rich portable devices such as Sharp's Viewcam® digital camcorders and Zaurus® PDAs.

The *joint value* of two complementary technologies $V(a_1 \& a_2)$ can be larger than the sum of two stand-alone values $V(a_1) + V(a_2)$ for two reasons. First, technologies developed in one firm may not be readily applicable to other firms who possess a different set of resources. Therefore, the stand-alone value $V(a_i)$ tend to be small due to implementation difficulties. Second, the innovating firm is in a better position to identify and promote synergy from the pool of internal technologies, hence enhancing the value beyond the simple sum of individual components. In the following analysis, I will use the phrase *internalized value* to describe the difference between the joint value $V(a_1 \& a_2)$ and the sum of the external values $V(a_1) + V(a_2)$.

Note that this is a very general framework. Internal complementarity does not have to involve a tangible integration process of concrete technologies. It may well arise from the corporate culture, routines, or organizational structure that make a technology more valuable internally than if used by other market players. Note that even within the same firm, the degree of internal dependence may vary across different types of knowledge.

Depending on the legal and social institutions, there is certain imitation risk p ($0 \leq p \leq 1$) in the economy, which firms take as given. With probability p, imitation would happen to technology a_i and the value $V(a_i)$ would be taken away from the firm. This is independent of what might happen to the other technology. Only when a_1 and a_2 are both imitated will firm *A* lose the whole value $V(a_1 \& a_2)$.[2]

An illustration of this scenario is Dupont's entry into the biotech field. The goal is to integrate biology with Dupont's current strengths and bring the production to a higher level. As Tom Connelly, Dupont's Senior Vice President and Chief Technology Officer put it, "new opportunities are going to come at the interfaces… For us, Integrated Science means … bringing on that additional capability and then looking for opportunities where more than one science comes together." Sorona®, the latest addition to the polymer platform, is a successful integration of new biological capabilities with Dupont polymer.

In this example, a_1 is the 3GT polymer with a series of desirable attributes, and a_2 is the biotech method to produce fiber-grade element for the polymer. Even if a_1 is imitated, Dupont can still stay competitive because, so far, the biotech method a_2 is the only commercially viable way to produce a_1. The same is true if only a_2 is imitated. Because of the proprietary 3GT polymer that a_2 is designed for, a_2 is much more valuable to Dupont than to any other firms.

3.3 Multinational R&D

In essence, there are two boundaries in multinational R&D: the *firm boundary* and the *national boundary*. Within a firm, complementary technologies create synergy in a way that is hard to duplicate outside the firm. Within a country, R&D activities are subject to a certain institutional environment that is distinct from that of other countries. MNEs are of particular interest because they expand their boundaries across multiple institutions, and create value from technologies that are exposed to different external environments.

Suppose there are two countries. Country *X* has strong IPR protection (small p_x), hence high price of human capital (large c_x); country *Y* is just the opposite: large p_y and small c_y.

[2] Even if both technologies are imitated, it is still uncertain whether the integration process is replicable. The development of these technologies involves such intensive interactions within the firm that the synergy is not readily transferable. Rivkin (2000) shows that interaction among the elements of a strategy makes imitation difficult. This factor is abstracted from here to keep the main theme traceable.

Firm *A*, with its two complementary projects, can now choose from the following three strategies,[3] as shown in Figure 1:

- Having both projects developed in country *X*;
- Having both projects developed in country *Y*;
- Having one project developed in country *X* and the other developed in country *Y*.

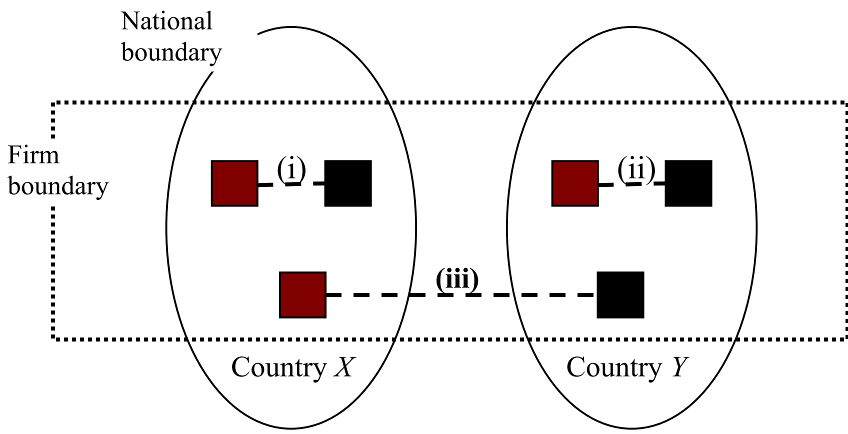


Figure 1. Organizational Strategies of R&D Projects

The market price of human capital in country j ($j = x$ or y) depends on the expected returns from innovation: the higher the imitation risk p, the smaller is the value appropriable by the innovators, and the lower is c_j. For individual firms, c_j is taken as given. Besides, cross border R&D incurs extra coordination costs R (Kuemmerle 1997).

Now we are ready to analyze the potential arbitrage opportunities of strategy (iii), in which the two complementary components are developed in two different environments with an institutional gap $\Delta p \equiv p_y - p_x > 0$ and a factor price difference $\Delta c \equiv c_x - c_y > 0$. [4]

First, compared with strategy (i), bringing a_2 to country *Y* means additional risk exposure. Not only is the stand-alone value $V(a_2)$ exposed to higher risks, the internalized value

[3] Firms can choose the organizational structures for their R&D even within one country. For example, R&D by small startups is common in the U.S., whereas large and diversified business groups dominate the R&D arena in many Asian countries. These organizational variances are assumed away in this model so as to highlight the cross-institutional implications; they are being studied in a separate project.

[4] Theoretically, the gap in factor prices will be endogenously affected by firms' arbitrage activities. However, this is beyond the scope of this paper, and should not be consequential unless we are looking into the very long run. The general equilibrium case is analyzed in a following project.

gets more risky too, since one of the two complementary components is now in country *Y*. However, with strong complementarity, $V(a_2)$ can be very small. The risk of the internalized value is also reduced if country *X* has strong IPR protection. The reason is that the internalized value stays within the firm as long as *either* of the two technologies remains exclusive. The interaction between firms and institutions comes into play in the presence of strong complementarity within the firms. Compensating for the extra risks, the firm gains from the less expensive human capital in country *Y*. The cost savings are large if a_2 needs a large amount of human capital to develop. Of course, the extra coordination cost *R* has to be justified by the net gains.

Second, compared with strategy (ii), keeping a_1 in country *X* means that the values from innovation are better protected. The stand-alone value $V(a_1)$ is safer, and the internalized value is subject to smaller imitation risks. The higher the imitation risk in country *Y*, the more advantageous it is to maintain part of R&D in country *X*. For purely domestic firms in country *Y*, even the internalized value is hard to protect because both complementary components are subject to the same weak IPR. The cost of the better protection is the higher R&D expenses on a_1, as well as the additional coordination cost *R*.

The cross-border arrangement would be a preferable strategy if the expected returns from strategy (iii) is larger than both (i) and (ii). The above discussion clearly indicates that the interaction between firm organization and external environment comes into play in the presence of internal complementarity.

3.5 Internalization-Arbitrage

I was faced with a puzzle: why do MNEs ever conduct R&D in countries with weak IPR protection? The answer involves an opportunity, and the ability to catch the opportunity.

Due to the poor IPR protection in some countries, local talents cannot efficiently realize value from their innovative activities. As a result, human capital is underutilized and underpriced. Such low cost human capital is attractive to R&D intensive firms if the firms possess alternative mechanisms to protect their intellectual property. One of the mechanisms, as identified above, is the substitution of *internal* organizations for external institutions. This allows firms to *arbitrage* the difference in factor prices across countries. I call it *internalization-arbitrage*.

There are two main factors involved in internalization-arbitrage.

First, different R&D projects are different in their internal dependence. For example, software development in India and product-based technical solutions in China are very human capital intensive, but the value of these technologies is highly dependent on the firms' overall R&D architecture. In contrast, projects developed in the home country, such as system designs or key interfaces, tend to be more valuable on their own. The strategy of differentiated project assignment, in essence, is to tailor and allocate projects so that the firm can capitalize on the strengths of particular locations while minimizing the costs and risks. In particular, firms can examine their R&D portfolios and allocate R&D projects with stronger internal dependence in weak IPR countries.

Exercising this strategy, however, is not as straightforward as it seems. Depending on the technology field and the firm's organizational structure, it can be a challenge to carve out the right projects for weak IPR countries. R&D projects in the IT industry are usually easier to decompose than in those the traditional industries, where the lack of a comprehensive knowledge base in the host country would significantly affect R&D efficiency and thus compromise the cost savings.

Second, MNEs, by keeping the key components under a strong IPR regime, are in a unique position to protect the internalized value arising from internal complementarities. With small imitation risk in the home country, having one project developed overseas brings little damage to the internalized value. Meanwhile, when the imitation risk in country *Y* is high, keeping the complementary components in the home country becomes critically important. Therefore, when the coordination costs are reasonably low, firms can actually leverage the home institutions for their operations overseas. The higher internalized value can also be considered higher leverage across countries.

The multilingual technology on computer systems is a good example of internalization-arbitrage. Because of the non-alphabetical nature of Chinese characters, Chinese software companies have been developing multilingual editors and related products since the early 1980s. These companies failed one after another due to the prevalence of software piracy. At the same time, Microsoft, IBM and Intel are actively developing multilingual technologies in China. The resulting achievements have significantly increased the international appeal of their products ranging from office applications to communication devices; the value-added is evidenced by their ever-increasing investment in this field.

In this example, value appropriation is not reliant on the firm-specificity of the R&D projects. Most local firms acknowledge the value of multilingual technologies and are able to integrate the technologies into their own products. What makes the difference is the inimitable component in the joint value: Intel chips, IBM business solutions, and Microsoft software packages marketed in strong IPR countries. While local firms have to face prevailing piracy and slim profit for their products, MNEs can appropriate the value of these innovations on the global market, where intellectual products are well protected and rewarded.

The example shows that internal complementarity can take on very general forms. The complementary component can be a technology, a firm-specific expertise, or simply the access to IPR-friendly markets. Firms can appropriate the value from R&D as long as the complementary components are not subject to the same high risks.

Of course, only those firms with sufficient organizational capabilities will find it worthwhile to set up R&D facilities in a foreign country, not to mention a country with very different institutional environment. The organizational capabilities differ widely across firms, depending on the their previous exposure to institutional idiosyncrasies, the experience in utilizing foreign technologies and resources, and the established routines of intra-firm knowledge transfers.

IV. EMPIRICAL DESIGN

The interviews and anecdotal stories are supportive of the internalization-arbitrage conjecture, but they are silent on how general the phenomenon is. In this step, I bring the idea to the U.S. patent data and construct empirical measures for the theoretical concepts.

4.1 Empirical Implications of Internalization-Arbitrage

The key insight that emerges from the qualitative analysis is the interaction between firms' internal organization and the external environments. An MNE can effectively conduct R&D in weak IPR countries if the value of the technologies can be retained inside the firm, and if the firm can efficiently coordinate cross-border activities. This lends a natural framework to the empirical analysis.

First, I focus on firms that conduct R&D in weak IPR countries and examine the *strategies* they use: how do firms arbitrage? If firms strategically allocate their R&D projects in response to the external environment, as suggested by the theoretical analysis, then we should be able to observe systematic differences among technologies developed under different IPR regimes. Specifically, technologies developed in weak IPR countries should have stronger internal linkages, controlling for firm characteristics.

Second, I compare across all firms and examine firm *capabilities*: who is doing the arbitrage? Internalization-arbitrage is a viable strategy only for firms with strong organizational capabilities. Hence, we should be able to observe systematic differences between firms that do R&D in weak IPR countries and those that stay away. Specifically, firms who are able to do so should have tighter organization of their innovative activities, controlling for locations.

Let γ_k be the degree of internal linkages for technology k, which is owned by firm i and developed in country j, then γ_k should depend on (1) the external IPR environment under which k is developed, and (2) the characteristics of the firm that develops the technology. Within firm i, γ_k is expected to be higher if j is a country with weak IPR protection.

Across firms, γ_k is expected to be higher if firm *i* has R&D in weak IPR countries, everything else being equal.

Before getting to the empirical tests, I need to (1) quantify the IPR regimes, (2) describe the data sources and the sample, and (3) construct variables from the sample data. I will address the tasks in detail in the next five subsections.

4.2 Weak IPR Countries

Arbitrage opportunities exist in a country that has a substantial reserve of human capital, yet suffers from weak IPR protection. Referring to the *World Development Indicators*, I exclude countries and areas that have under two million population, less than 1% of gross tertiary school enrollment rate, or less than five patents filed with U.S. Patent and Trademark Office (USPTO) in the 1990s.[5] The reason for the screening is that a small human capital reserve can hardly justify multinational entries. The few inventors who do reside in these countries are very likely to represent extreme cases, which may cause biases rather than adding explanatory powers. The removed countries and areas are, for example, Belize, Trinidad, and most sub-Sahara African countries. For the same reason, I remove war-torn countries such as Croatia and Yugoslavia.

Six indices are considered to measure the institutional environment for IPR protection; they reflect different aspects of institutional environment and have been widely used in literature. The time horizons covered by these indices do not exactly coincide with each other. However, given the slow changes in institutions, these indices should be indicative of the IPR environments in these countries.

The first three indices apply to the general legal and political environment:

- The Law and Order index in the *International Country Risk Guide* (*ICRG*) *Risk Rating System* (1993-97). The index is formed using public sources such as newspaper reports published in the country in question, national and international news services, reports of

[5] When eliminating countries with negligible innovations, I also cross-referenced World Intellectual Property Organization (WIPO) data for residents filed patents.

national, regional and trans-regional banks and other institutions, and international organizations such as OECD, BIS, IMF and the World Bank.

- The O-Factor in the *PricewaterhouseCoopers Opacity Survey* (2000). Opacity is "the lack of clear, accurate, formal, easily discernible, and widely accepted practices." The potential for opacity exists in five principal areas: corruption in government, the laws governing contracts or property rights, economic policies, accounting standards, and business regulations. A high degree of opacity in any of these categories will expose the appropriable value of R&D to higher risks.
- The Property Protection index in the *Index of Economic Freedom* (1995) by the Heritage Foundation. It mainly tracks seven aspects of property rights protection: the commercial code defining contracts, sanctioning of foreign arbitration of contract disputes, government expropriation of property, corruption within the judiciary, delays in receiving judicial decisions, and legally granted and protected private property.

The second set of indices apply specifically to intellectual property rights protection:

- Rapp and Rozek (1990) index. This index reflects the conformity of national patent laws with the minimum standards proposed by the US. Chamber of Commerce. It covers about 97 countries, and pertains to the situation in the mid-1980s. Outdated as it might be, this index is still a valuable reference on IPR environments for most countries.
- Ginarte and Park (1997) index. This frequently cited index was produced for five-year intervals starting in 1960 and ending in 1995. I use the data covering the most recent period: 1990-95. The index rates the national patent protection system according to five categories: the extent of coverage, membership in international treaties, provisions for loss of protection, enforcement mechanism, and duration of protection.
- United States Trade Representative's Special 301 watch list and priority watch list (1999). On the lists are "trading partners that deny adequate and effective protection of intellectual property or deny fair and equitable market access to United States artists and industries that rely upon intellectual property protection."

I also supplement the indices with the Rule of Law index developed by Kaufmann, Kraay and Zoido-Lobatón (1999, 2002), and with the piracy index developed by the International Planning and Research (IPR) on behalf of the Business Software Alliance (BSA) and the Software & Information Industry Association (SIIA).

These indices turn out to be quite consistent. Whether I use each single index, or a composite index with various weights, I get a reasonably stable list of 34 countries with weak IPR protection. The country names and the corresponding indices are listed in Appendix A. Interestingly, this list is not restricted to low-income regions. The per capita GNI ranges from U.S.$25,920 for Hong Kong (China) to U.S.$440 for Pakistan in 2000, according to the World Bank statistics.

4.3 Patent Data

Despite the various criticisms of patent data, I choose to use the U.S. patent data for the following reasons (in addition to Griliches 1990; Patel and Pavitt 1995).

First, to meet the criteria for patenting, a technology has to be novel, non-obvious, and useful. Therefore, using patent data allows me to eliminate the localization/adaptation type of R&D specific to the host countries, and instead focus on overseas innovations that can bring value to the whole firm – the home base augmentation type of R&D as defined in Kuemmerle (1999). Moreover, because patents are the output of R&D, they capture the projects that *fruitfully* utilize human capital in various countries. [6]

Second, patent citation is one of the most traceable evidences of knowledge flows (Jaffe et al. 2000). The systematic documentation of patent citations tracks the knowledge flows within and across the firms' global innovation networks. For example, the citations received by each patent indicate who are following up on the invention, when and where.

Finally, the detailed location information for patent inventors can help me identify the geographic distribution of talents utilized by the U.S. firms, as well as the collaboration

[6] Also under consideration is the publication data for basic research (versus patents data for technology development). However, propensity to publication may be even more difficult to control for.

among them. Since the inventors' mailing addresses are required for patent applications, they should reliably reflect the actual locations where the innovations take place.

Admittedly, the usual caveats apply: not all innovations obtain patents, and not all knowledge flows are reflected in patent citations. Casual observations suggest that most R&D results in weak IPR countries are not patentable. In addition, a significant proportion of citations are imposed by patent examiners rather than voluntarily provided by the applicants. It would be unlikely that those imposed citations represent any systematic learning. I will analyze these potential biases in detail in later sections.

Information on patents granted between January 1993 and December 1999 is obtained from the NBER patent data (Hall et al. 2001). Patents granted between January 2000 and August 2003 are extracted from the Grant Red Book V2.5 bibliographic data of USPTO. Every field is closely examined to ensure consistency when the two datasets are merged.

Since patents may be assigned to parent companies or their subsidiaries for unobservable reasons, I study each multi-unit firm as an integrated strategic agent. The Directory of Corporate Affiliation (DCA) database published by Lexis-Nexis traces corporate linkages of more than 174,000 parent companies, affiliates, subsidiaries, and divisions worldwide. It allows me to build family trees for each *firm* in my sample. An American firm in this study refers to an ultimate parent registered in the United States, plus all its subsidiaries and affiliates, home as well as abroad, that are owned by the same ultimate parent, directly or indirectly, by more than 10%.[7] After aggregation, the number of firms in my sample is only about half the number of assignees.

The Industrial Compustat data by Standard & Poor's are used to capture other firm characteristics, such as industries, assets and sales. To avoid noise from small or instable firms, I drop companies with less than $0.1 million assets, as well as those that are traded for less than three years in the sample period. It is true that using Compustat data limits the study to publicly traded companies, but it also eliminates the incomparability problem

[7] This is the official criterion of FDI for U.S. companies. I also varied the ownership threshold for robustness checks and found no substantial change in results.

between public and private firms, as these two types of firms are subject to very different operational constraints. Given the fact that public firms hold the majority of U.S. patents (Hall et al. 2001), they should be representative of innovating firms.

4.4 The Sample

I decide to focus on U.S.-headquartered firms in the empirical study. The reason is that I can obtain the most comprehensive data for U.S. firms, and at the same time avoid the potential administrative biases in cross-country comparisons.

The sample period is from 1993 to 2001, and I study all the patents applied during this period (and granted up to August 2003). Due to a typical lag of 2-3 years between patent application and patent granting, including the recently applied patents will bias the sample toward "quick patents," which may have systematically different characteristics. At the same time, R&D in weak IPR countries is still a recent phenomenon. Extending the sample period further back will not add much value to the analysis either.[8]

Industries vary widely in their propensity to patenting and the usefulness of patents as the measure of innovative activities (Cohen, Nelson and Walsh 2000). The industry controls can only partially alleviate these problems. To reduce unrelated noises, I remove industries where patents are a very weak indicator of innovations (e.g., insurance), industries that are heavily influenced by public policies (e.g., utilities), industries that are mostly domestic (e.g., retailing), or industries whose geographic locations are dictated by some exogenous factors (e.g., mining). The main sample includes the following two-digit SIC industries: 28 (chemical and allied products), 29 (petroleum and coal products), 30-39 (other manufacturing), 48 (communications), 73 (business services, including software), and 87 (engineering and management services). I vary the industry selection in robustness checks and make sure that the findings do not depend on specific industries.

The challenge in the data preparation is to match the large datasets together, where the only link among them is the company names. I decide not to use the match offered in the

[8] As a robustness check, I also examine the sub-period from 1993 to 1997 in order to stay away from the irrational expansion during the Internet Bubble.

NBER patent data, which uses the 1989 Compustat data. Instead, matching is conducted year by year to accommodate possible organizational changes, which are not unusual during this period. An elaborate computer program[9] was developed for this purpose, and all the results are manually checked. Ambiguous matches are further verified via Dun & Bradstreet Million Dollar Database, company websites and industry publications.

The whole data cleaning procedure is illustrated in Appendix B. First, listed companies in the selected industries are taken from Compustat and then matched with the parent companies in DCA. For each matched firm, all the family members are extracted from DCA. Next, the parents, branches, subsidiaries, and affiliates are matched with the assignees of patents applied in the corresponding year. Assignee names are standardized according to the USPTO company name files. Finally, the patent information is aggregated at the parent level and merged with the Compustat-DCA match.

I drop those firms that do not have any patents during the entire sample period and those three-digit SIC industries that contain less than three innovating firms. After data cleaning, the main sample consists of 1,567 firms in 92 three-digit SIC industries, whose patent output during the nine-year period ranges from one to more than 20,000 (IBM), averaging at more than 100 per firm. This is a very skewed sample in the sense that the median firm only has nine patents, and that the top 25% firms filed nearly 95% of the patents in the sample period. Among these firms, 681 firms register positive utilization of foreign inventors while only 227 of them use inventors from weak IPR countries.

From these data sources, I am able to construct a rich set of variables for the empirical analysis. The key variables include the geographic distribution of a firm's innovations and the internal linkages among these innovations.

4.5 Location

Empirical studies using patent data often take the country of the first author as the location of the invention, partly because this information is readily available. However,

[9] I thank Wilbur Chung for his help with the program.

given my focus on human capital utilization, I want to be very careful about possible distortions. In a patent with multiple inventors, which is often the case in cross-border R&D projects, a foreign inventor is likely to be the second, third, or fifteenth coauthor in the sequence. The first author's location may therefore bias against the involvement of foreign inventors, particularly those from developing countries.

To verify the possible bias, I track the addresses of all the inventors whose names are listed on the patent applications, and weigh the contribution of each inventor by his/her sequence in the inventor list.[10] Then, for every patent, I calculate the percentage contribution from different countries. Comparison between the first-author rule and the weighted-contribution rule confirms my intuition that inventors in weak IPR countries are more likely to be collaborating with inventors from other countries, and they are more likely to be at the tail of the inventor sequences.

In the following analysis, I apply equal weights to all inventors regardless of sequence. If more than half of the inventors are from weak IPR countries, then the patent is considered to be developed in weak IPR countries. If more than half of the inventors are from the U.S., then this patent is considered to be developed in the home country. The rest are patents developed in other (strong IPR) foreign countries.[11]

Accordingly, a firm is considered to have R&D in weak IPR countries if at least one of its patents is developed in those countries. The firm is considered to have foreign R&D if at least one of its patents is developed in a foreign country. It turns out that the classification criteria have important implications to the analysis, although my results remain robust across specifications. I will discuss this point in the robustness analysis.

The 227,034 patents in the full sample are described in Table 1. The share of patents developed in weak IPR countries, small as it still is, has been steadily increasing over the last decade. Table 2 presents summary data for the 1,567 firms included in the sample.

[10] There are no absolute rules as to the weighting, although higher weights are usually assigned to inventors listed at the beginning of the sequence.

[11] For example, the Intel patent 5,677,862 has three inventors from Haifa, Israel and two from California, U.S.. Hence, it is considered a technology from Israel. This patent was later cited by another Intel patent 6,470,370, which is considered a US-developed technology as three out of five inventors are from U.S..

4.6 Internal Linkages

There is no direct measure of internalized value, but value can be proxied by usage. Technologies whose values are highly dependent on internal resources are more likely to be utilized within the firm. Since "citations to patents that belong to the same firm" represent internalized knowledge transfers "leading to competitive advantage" (Hall et al. 2001; 2003), I use self-citations to proxy for the internalized value of each technology. Presumably, the more a patent is cited by the same firm, the more its value is being retained inside the firm boundary.

Note that my measure of "self-citation" differs from that of Hall et al. (2001) in two important aspects, which explain the relatively higher self-citation ratios in this study.

First, HJT uses the patent assignee code as their unit of analysis. They acknowledge that "the same firm may appear in different patent documents under various, slightly different names," hence assuming different assignee codes[12]. Hence, "Dell USA Corporation" and "Dell USA, L.P." were treated as two different entities. To avoid this problem, the assignees in my sample are all matched to the DCA data of the corresponding year to make sure that every firm is unique and identifiable.

Second, HJT treats every assignee as an independent entity. Affiliates or subsidiaries of the same firm will be given their own assignee codes and hence will show up as unrelated. Because I am more interested in the firm as an integrated organization, any citations that occur among affiliated entities are considered self-citations. For example, a citation from Lotus Development Co. in Korea to IBM in Armonk, New York would be counted as self-citation if the patents were developed after the acquisition in 1995.

Similar to HJT, my citation calculation is subject to the truncation problem in the time series. The patents filed in the 1993-2001 sample period had only received a fraction of the citations by the end of August 2003. However, I believe that this problem will not significantly affect my results, as patents in the same firm or industry tend to be affected

[12] See footnote 22 of Hall et al. (2001).

similarly and I will certainly control for the between-group variations in my analysis. Even if there are significant within-group variations in citation lags, this measure is still consistent with my objective of capturing the efficiency in internal knowledge utilization: *speed* as well as *scale*. A self-citation ratio is high either because the technology is used more internally, or because the internal inventors are able to build on the technology faster, before the external citations take place.

V. EMPIRICAL RESULTS

In this section, I seek empirical evidence of the internalization-arbitrage conjecture by analyzing the within- and across-firm variations in the technology structures.

5.1 The Econometric Model

Let N_k be the number of citations received by patent k, among which n_k are self-citations. As discussed previously, the self-citation ratio n_k/N_k, which proxies for internal linkages, depends on the external IPR environment as well as the internal firm characteristics.

Econometrically, I choose the zero-inflated negative binomial (ZINB) model to reflect three features of the data.

First, most patents received only a small number of citations during the short sample period. A count model is more appropriate than a continuous one when 75% of the patents received fewer than five citations and two self-citations (Hausman et al. 1984).

Second, a negative binomial (NB) model is preferred to a simple Poisson model due to the large variance in the number of citations and self-citations received by each patent. In the following analysis, I fit both Poisson and NB models for every setup, although only the results from the NB model are reported. The over-dispersion parameter α is significantly different from zero in almost all setups (χ^2 = 4,700 and $pr > \chi^2$ = 0.000 in the baseline model), *conditional on* the independent variables.

Third, the frequent occurrence of zero self-citations may well arise from a different mechanism. Because many patents have yet to receive any citations by the end of the sample period, $n_k = 0$ does not necessarily mean a low level of internalization. It may simply be constrained by a small N_k. The Vuong statistic (Vuong 1989) for non-nested models shows large positive values (z = 12.49 and $pr > z$ = 0.000 in the baseline model) favoring the ZINB model versus the standard NB model.

In the first step, I focus on the 227 firms that conduct R&D in weak IPR countries and study the within-firm variations. The regression takes the following form:

Regime 1: $E(n_k) = N_k \cdot exp\ (\beta_0 + \beta_1 \cdot foreign + \beta_2 \cdot weak + \beta_3 \cdot i.firm + \Lambda_k)$

Regime 2: $E(n_k) = 0$

$Prob$ (Regime 1) = $exp\ (\gamma_0 + \gamma_1 \cdot N_k)/[1 + exp\ (\gamma_0 + \gamma_1 \cdot N_k)]$

Here *foreign* and *weak* are two dummy variables to indicate, respectively, whether the technology is developed in a foreign country and a weak IPR country. Hence, *foreign* must be 1 if *weak* = 1. *i.firm* represents 226 dummy variables for firm fixed effects. The exposure variable N_k serves as the scope in which self-citations can be observed for each patent. Finally, Λ_k represents a list of control variables such as application years and patent classes. A cluster model is used to allow the possibility that the observations are independent across firms but not necessarily within firms.

The coefficient on the variable *foreign* is expected to be negative. Previous studies have shown that knowledge diffusion is geographically concentrated in nature (Almeida 1996, Jaffe and Trajtenberg 1998). Cross-country knowledge transfer is difficult even within the firm (Teece 1977). Therefore, I expect the foreign developed patents to be less intertwined with the parent company's knowledge base.

Meanwhile, I expect a positive coefficient on the variable *weak*. The purpose of R&D in weak IPR countries is to tap the underutilized human capital, and internalization is used as a barrier against imitation. Hence, the internal linkages would be stronger in countries with weak external institutions.

The strong internal linkages may be due to the fact that firms allocate the intrinsically more internalized technologies in weak IPR countries. Alternatively, firms may develop similar technologies at multiple locations, but with different organizational structures. Supply factors may also play a role here, in the sense that the expertise available in the weak IPR countries happens to fit the needs of highly internalized R&D. These cases are all consistent with the theoretical conjecture. Although a control for technology fields is not necessary from the theoretical point of view, it would be interesting to see empirically whether it is the type of technologies that is driving the differences.

In the second step, I compare across all the firms in my sample and examine whether firms doing R&D in weak IPR countries manifest stronger internal linkages. The regression takes the form:

Regime 1: $E(n_k) = N_k \, exp\,(\beta_0 + \beta_1 \cdot f\_foreign + \beta_2 \cdot f\_weak + \beta_3 \cdot foreign + \beta_4 \cdot weak + \Lambda_k)$

Regime 2: $E(n_k) = 0$

$Prob$ (Regime 1) $= exp\,(\gamma_0 + \gamma_1 \cdot N_k)/[1 + exp\,(\gamma_0 + \gamma_1 \cdot N_k)]$

Here *foreign* and *weak* are the same as defined in the within-firm analysis. *f_foreign* and *f_weak* are two dummy variables to indicate whether the firm that owns patent *k* has any patents developed in foreign countries and in weak IPR countries. Hence, *f_foreign* must be 1 if *foreign* = 1 or *f_weak* = 1, and *f_weak* must be 1 if *weak* =1. Λ_k represents other control variables such as firm sizes and patent portfolios. Again, a cluster model is used to allow the possibility that the observations are not independent within firms.

According to the discussion in Sections III, firms need to have strong organizational capabilities to implement successful internalization across borders, even more so if they want to conduct R&D in weak IPR countries. Therefore, I expect positive coefficients on both *f_weak* and *f_foreign*.

The variables *weak* and *foreign* remain in the regression because locations significantly affect the degree of internalization. For example, a patent developed by IBM in Japan would probably have a lower self-citation ratio than a patent developed by Motorola in the U.S.. However, this comparison does not say much about the R&D organizations of

IBM and Motorola. By controlling for the location effect, I am able to examine whether, among the technologies developed under similar environment, a firm's R&D presence in weak IPR countries is associated with stronger internal linkages among its technologies.

Again, the difference in internal linkages across firms may due to their heterogeneous organizational capabilities. It may also arise from the particular technology fields that the firms are involved in. I cannot precisely differentiate these two cases, as the technology fields may be endogenously chosen by firms with different capabilities. For the purpose of this study, it is sufficient to identify the relationship between a firm's internal technology structure and its multinational R&D strategies.

In both models, the marginal effect (economic significance) of the independent variables can be calculated as follows:

$$E(n_k) = N_k \exp(\boldsymbol{\beta}'\mathbf{x}) \;\Rightarrow\; \frac{\partial E(n_k)/N_k}{\partial x_i} = \frac{E(n_k)}{N_k} \cdot \beta_i$$

When the independent variables are dummy variables, the coefficients can be roughly interpreted as the percentage change in the self-citation ratio if the corresponding independent variable changes from 0 to 1:

$$\beta_i = \frac{\partial[E(n_k)/N_k]}{E(n_k)/N_k}$$

5.2 Within Firms: Differentiated Project Assignment

In this step, I look into the 227 firms that conduct R&D in weak IPR countries, and compare the technologies they develop under different IPR regimes. Table 3 gives the average self-citation ratios for three groups of patents: those developed in weak IPR countries, those developed in strong IPR foreign countries and those developed in the home country (U.S.). As expected, the first group consistently shows higher self-citation ratios than the second, highlighting the effect of external environment. The increasing self-citation ratios over time simply reflect the fact that self-citations happen faster than citations across organizations, and this difference becomes more salient when the observation window gets narrower.

The regression results are shown in Table 4. Column (1) is the baseline model with 226 firm dummies and eight year-dummies to remove the time effect. Replacing the year-dummies with a trend variable generates similar results. In column (2), I only include firms that have more than 50 patents over the nine-year period. The purpose is to make sure that the results are not driven by misrepresentative observations in small firms. Since the theoretical conjecture is on firms' internal complementarities, firms above certain R&D scales should be the more appropriate group to study.

Statistics show that over half of the citations by US patents are imposed by examiners, rather than voluntarily filed in patent applications (Alcacer and Gittelman 2003).[13] Although inventors may fail to cite the prior arts for various reasons, it would be far-fetched to interpret the examiner-imposed citations as knowledge flows. Luckily, the distinction between inventor-filed and examiner-imposed citations is now available for patents granted after 2000. In column (3), I conduct the test using only voluntary citations made by post-2000 patents.

In the theoretical analysis, I argue that the viability of the internalization-arbitrage strategy may vary widely across industries. It would be informative to examine not only the aggregate, but also the particular technology fields. In Column (4), I control for patent classes[14] to test whether the difference in internal linkages are driven purely by the distribution of technology fields across countries. Column (5) reports the results for Computers and Communications,[15] an industry I conducted most of my interviews in. Separate tests are also conducted in other technology categories.

Throughout the specifications the coefficient on the *foreign* dummy is significantly negative, which confirms the distance effect. Within the same firm, foreign developed patents are 20-30% less likely to be cited by the same firm, compared with those developed in the home country. The difference is even larger in the computer industry.

[13] I thank Juan Alcacer and Michelle Gittelman for their help on data preparation.

[14] The reported result uses the 3-digit U.S. primary patent classes. Fully aware of the caveats with this classification, I also use the International Patent Classes (IPC) for robustness tests.

[15] As defined in Hall et al. (2001), this category covers 35 primary patent classes spanning communications, computer hardware & software, computer peripherals, and information storage.

However, the positive coefficient on the *weak* dummy nearly offsets all the negative distance effect. This is not trivial if we believe that the weak IPR countries are even further away from the home country in terms of culture, institutions and technological development. The net effect of *weak* and *foreign* indicates that technologies developed in weak IPR countries are intertwined in the firms' internal knowledge base as if they were right at U.S. headquarters!

It is worth pointing out that the regression results vary widely across technology fields. Generally speaking, the coefficients are economically and statistically significant, with the expected signs, in IT, biotechnology, and other "new" industries, but not in traditional areas such as chemical and machinery. This confirms the earlier discussion on the difficulty of decomposing R&D projects and the importance of a broad knowledge base in certain industries.

Finally, I cannot rule out the possibility that the variance in the self-citation ratios is simply driven by the development level of the host countries. Most of the weak IPR countries are also less developed countries. Innovators in those countries have to rely heavily on the knowledge base of the parent companies and therefore produce very firm-specific intellectual properties. This interpretation, however, is still consistent with the theoretical argument. MNEs provide the complementary knowledge that the local talents cannot obtain otherwise. As a result, the generated technologies are less applicable outside the firm boundaries. The low absorptive capacity in the local community does not seem to be the explanation. In fact, the results remain strong even if I only study citations made from U.S.-developed patents.

5.3 Across Firms: Organizational Structure

In this step, I compare across firms and examine firm capabilities. Do firms that conduct R&D in weak IPR countries differ systematically from others regarding their knowledge organization? Table 5 gives the average self-citation ratios for three groups of firms: those with positive patent output in weak IPR countries, those with positive patent output in foreign – but not weak IPR – countries, and those whose R&D is in the home country

(U.S.) only. Statistics for large R&D firms are also shown in the table. Consistently, firms with R&D in weak IPR countries show much higher self-citation ratios.

The regression results are shown in Table 6. Column (1) is the baseline model on the full sample, with fixed year effect. Column (2) constrains the sample to firms with more than 50 patents during the nine-year period. Same as in the within-firm analysis, column (3) removes examiner-imposed citations and limits the sample to voluntary citations received after January 2000. Patent classes are controlled for in Column (4). Column (5) gives the results for the field of Computer and Telecommunications.

Across specifications, the coefficients on *f_weak* and *f_foreign* are positive and significant. The results show that firms who conduct R&D overseas generally have more internalized technology structures compared with purely domestic firms. In addition, among firms with foreign R&D, those who do R&D in weak IPR countries manifest even stronger internal linkages. Since location effects are controlled for throughout the analysis, the results are not driven by the geographic distribution of firms' patent portfolios.

Arguably, firms with R&D in weak IPR countries have proportionally more self-citations simply because they are larger firms. I would like to examine more closely whether the degree of internalization is fully explained by firm size. In Table 7, I report the regression results controlling for assets (1), sales (2), and the total number of patents owned by the firm (3). I prefer not to include these controls in the main models because it is still consistent with my argument that only large firms have the complementary resources for internalization-arbitrage.[16]

Surprisingly, assets and sales both show up with negative coefficients. That is, among firms that do R&D in foreign/weak IPR countries, larger size is associated with even weaker internal linkages, although the marginal effect is almost negligible. Meanwhile, the coefficients on *f_weak* and *f_foreign* remain positive and significant. A large patent pool does offer some explanation power to the high degree of self-citations, but the economic significance is, on average, much smaller than that of *f_weak* and *f_foreign*.

[16] An alternative argument is that, optimally, firms expand until the organizational cost goes above the benefit of internalization. In that sense, firm size itself reflects organizational capabilities.

Similar results are obtained when I use the logarithms of these size variables. Therefore, the data provide supportive evidence that firms with R&D in weak IPR countries feature stronger internal linkages, even after controlling for firm sizes.

VI. ROBUSTNESS ANALYSIS

In this section, I discuss the caveats in the above analysis and carry out a series of robustness tests to make sure that the results do not depend on the specific setups.

6.1 Bias in Using Patent Data

In the absence of a direct measure for innovation, using patent data may generate biases in both steps of the empirical analysis. Within firms, I hope to test whether *technologies* developed in weak IPR countries are used more internally. Instead, I can only observe whether *patents* developed there receive proportionally more self-citations. Across firms, I hope to test whether firms that *do R&D* in weak IPR countries have tighter internal structures. Instead, I can only observe whether firms that *have patents* from there are more internalized. With different patent propensities across firms and locations, I need to examine to what extent the unobservable part of R&D activities would affect my results.

Firms file patents so that they can "exclude others from making, using, or selling the technology in the United States" (USPTO), and claim credit from whoever uses it. Given the nontrivial cost involved in patent applications, firm will only patent a technology if it expects potential uses of the technology by other firms. This is true even when patenting is purely for strategic reasons (Hall and Ziedonis 2001). Therefore, patents should reflect the part of R&D that is relatively less internalized. Meanwhile, the results of many R&D activities, such as software codes and laboratory tests, are not readily patentable. These activities also tend to be the more firm specific. As a result, self-citation ratios calculated from the patent data would underestimate the overall internal linkages of R&D activities.

Conversations with managers and engineers in the multinational R&D labs in China suggest that most R&D efforts there are not in the patent-oriented category. Only a small proportion of the technologies with the highest stand-alone value will be patented in the U.S.. If non-patentable R&D is more common in the weak IPR countries – which seems to be the case according to observations – then the estimation of internal linkages is biased downward to a larger extent in weak IPR countries than in other countries. If anything, this sample bias goes against my findings and should make the results even more significant. The same logic applies to the analysis of firms.

6.2 Biases in Measuring Self-Citations

Due to the changes in firm organizations over the years, the self-citation counts may be biased upward or downward. For example, after firm *A* is acquired by firm *B*, *A* as a firm no longer exists. As a result, any citation from the post-acquisition firm *B* to the prior arts of *A* would not be counted as self-citations, even if the citing and the cited patents involve exactly the same team of inventors. On the other hand, if a team in firm *A* worked on a project before the acquisition date and the resultant patent is assigned to the post-acquisition firm *B*, then citations to the prior arts owned by *B* would be counted as self-citation, although they hardly represent any within-firm knowledge flows.

To avoid these biases, I eliminate from the sample any assignees that changed their firm affiliations during the sample period. Also eliminated are firms that substantially changed their names, even though the affiliation codes remain the same. This decreased the number of firms in my sample to 1,054.

The same within-firm and across-firm regressions are conducted on the reduced sample, and the results for the two baseline models are reported in Table 8. Compared with the results in Table 4 and Table 6, the reduced sample with stable firm affiliations produces even stronger within- and across-firm variations. Controlling for firm effects, patents developed in weak IPR countries are 25% more likely to be cited internally than those developed in other foreign countries. Controlling for locations, firms with R&D in weak IPR countries have about 43% higher self-citation ratios than firms doing R&D in strong IPR countries, and 74% higher than the purely domestic firms.

6.3 Alternative Measure of Internal Linkages

There are many forms of internal linkages for an organization, and the self-citation variable is at best an imperfect proxy. The internalization-arbitrage argument would be more convincing if the same pattern can be identified with alternative measures.

One measure that has been developed for this purpose is R&D collaboration across geographic locations (Lahiri 2003). Presumably, having researchers from different countries collaborate on the same project not only signals the firm's strong coordination capabilities, but also facilitates future knowledge flows within the firm. Suppose that in a particular firm, the number of patents that involve inventors from country c is K_c. Among them, k_c patents are the results of collaboration with inventors from other countries. Hence, the ratio k_c/K_c captures the linkages between the inventors in the country c subsidiary and other parts of the firm.[17]

Again, I follow the two-step procedures of within- and across-firm analysis. Focusing on firms with R&D in weak IPR countries, I find that on average, 62% of the patents with inventors from weak IPR countries are the results of multinational collaboration. The number is 44% for patents with inventors from other foreign countries. Controlling for the same location, say, the strong IPR foreign countries, the k_c/K_c ratio is 44% for firms conducting R&D in weak IPR countries, and is 37% for others.

This calculation, therefore, confirms the thought that technologies developed in weak IPR countries feature stronger internal linkages, and that the firms tapping talents in weak IPR countries consistently have stronger internal linkages.

VII. CONCLUSION AND FUTURE EXTENSIONS

Doing R&D in countries with weak IPR protection appears to contradict the comparative advantage theory. With poor institutional environment, these countries are not known for their R&D or technology strengths. This study is an attempt to address the puzzle.

[17] In fact, the collaborated patents have, on average, 20% higher self-citation ratios than patents whose inventors are from the same country. This difference alone merits further exploration.

The interviews, the theoretical analysis, and the empirical evidence consistently point to one potential explanation: MNEs may be substituting their internal innovation organizations for the external environment. Thus, firms with closely-knit internal knowledge structures are able to take advantage of the underutilized human capital in weak IPR countries without exposing themselves to too much risk.

This adds to our understanding of a fundamental question in international strategy. Traditional views focus on firm-specific assets (intangibles, etc.) or location-specific endowment (natural resource, etc.) as the driving forces of internationalization. More recently, Ghemawat (2003) advocates that the foundation for international strategy should be built on the arbitraging of international differences, strategies made possible by internal firm capabilities. This study is a direct illustration of this point: institutional gaps across countries can be an important source of arbitrage opportunities. The study also reveals the internal capabilities required to take advantage of the opportunities. Just as globalization is not for everybody, neither is setting up R&D centers in China or India. To benefit from R&D internationalization, a firm's internal infrastructure matters.

The analysis also has important implications to the local firms in weak IPR countries. To survive the adverse institutional environment, and to compete effectively with MNEs, the local firms have the alternative of expanding beyond national boundaries and leveraging foreign institutions for value appropriation. In fact, we have already seen this possibility in a few Indian high-tech firms, although the organizational costs are still intimidating for most firms from emerging economies.

For policy makers, this study suggests that MNEs are now able to respond to national policies with global strategies. Firms can leverage one country's institutional environment for their operations in other countries, and public policies in one country may have spillover effects on the effectiveness of other countries' policies. Therefore, policy makers may have to keep a more dynamic view of the problems they are facing.

While identifying the potential opportunities in the adverse environments, I am by no means indicating that poor IPR protection is good either for firms or for the economy. In face of weak legal institutions, firms have to strategically internalize their knowledge-intensive activities, and only a small number of firms are able to do so. In other words, internalization-arbitrage is a strategic choice under the constraint of poor external environment. Removing this constraint is expected to improve both the channel and the direction of knowledge flows. For example, MNEs would feel more comfortable transferring technologies that are more readily applicable to the host countries. Indeed, Branstetter et al. (2002) show that U.S. MNEs responded to IPR reforms in the host countries with more technology transfers to the subsidiaries.

This paper also opens a whole line of further research.

First, the essential message in this study is that sustainable advantage does not come from the most favorable environment, but from the right interaction between firm capabilities and external idiosyncrasies. Since the external environment varies along many dimensions such as competition, technology changes and government regulations, there are many interesting questions worth exploring. For example, do firms conduct a specific subset of R&D at technology clusters, where they face close interaction with competitors? What is the relationship between this subset and the technologies they develop elsewhere? How does this relationship affect firms' collocation decisions?

Second, efficient coordination is shown to be a crucial condition for synergy. Several alternative measures have been developed in this study to proxy for firms' internal coordination. In the next step, I would like to examine the effect of coordination efficiency on firms' diversification behaviors, the benefit from diversification, and their overall performance.

Third, it would be interesting to take the current analysis to a dynamic setting. Besides information technologies, what were the fundamental changes that triggered the start of internalization-arbitrage? Are these changes in corporate management or the external environment? Or both? Or is one the repercussion of the other?

Finally, the arbitrage-internalization mechanism should not be limited to innovation only. Similar arguments can also be found in the mainstream internalization theories (Buckley and Casson 1976) and in the line of work on "business groups" (Khanna 2000). While the intangibility of knowledge makes R&D the ideal field for institutional arbitrage, we have reason to believe that the mechanism applies to other fields too. A study with a broader view would surely deepen our understanding of firms and institutions.

REFERENCES

Alcacer, Juan and Michelle Gittelman. 2003. Patent Citations by Inventors and by Examiners: the Implications to Knowledge Flows. Working Paper.

Almeida, Paul. 1996. Knowledge Sourcing by Foreign Multinationals: Patent Citation Analysis in the U.S. Semiconductor Industry. Strategic Management Journal, 17(Special Issue: Knowledge and the Firm, Winter): 155-65.

Anand, Bharat and Alexander Galetovic. 2003. Strategies That Work When Property Rights Don't. Advances in the Study of Entrepreneurship, Innovation, and Economic Growth, JAI Press.

Bartlett C. A. and Ghoshal S. 1990. The Multinational Corporation as an Interorganizational Network. Academy of Management Review, 15(4): 603-625.

Branstetter, Lee, Ray Fisman and C. Fritz Foley. 2002. Do Stronger Intellectual Property Rights Increase International Technology Transfer? Empirical Evidence from U.S. Firm-Level Panel Data. Working paper.

Buckley, Peter J. and Mark Casson, C. 1976. The Future of the Multinational Enterprise, London: Palgrave-Macmillan.

Cantwell, John and Odile Janne. 1999. Technological Globalisation and Innovative Centres: the Role of Corporate Technological Leadership and Locational Hierarchy. Research Policy, 28(2-3): 119-44.

Chung, Wilbur and Juan Alcacer. 2002. Knowledge Seeking and Location Choice of Foreign Direct Investment in the United States. Management Science, 48(12): 1534-54.

Cockburn, Iain and Rebecca Henderson. 1998. Absorptive Capacity, Coauthoring Behavior, and the Organization of Research in Drug Discovery. Journal of Industrial Economics, 46(2): 157-182.

Cohen, Wesley M. and Daniel Levinthal. 1990. Absorptive Capacity: A New Perspective on Learning and Innovation. Administrative Science Quarterly. 35: 128-152.

Cohen, Wesley M. and Steven Klepper. 1996. A Reprise of Size and R&D. The Economic Journal, 106(437): 925-51.

Cohen, Wesley M., Richard R. Nelson and John P. Walsh. 2000. Protecting Their Intellectual Assets: Appropriability Conditions and Why U.S. Manufacturing Firms Patent (or Not). NBER Working Paper 7552.

Frost, Tony S. 2001. The Geographic Sources of Foreign Subsidiaries' Innovations. Strategic Management Journal, 22(2): 101-23.

Ginarte, Juan Carlos and Walter G. Park. 1997. Determinants of Patent Rights: A Cross-National Study. Research Policy, 26: 283-301.

Ghemawat, Pankaj. 2003. Semiglobalization and International Business Strategy. Journal of International Business Studies, 34(2): 138-152.

Greene, William H. 2002. Econometric Analysis. 5th Edition. New Jersey: Prentice Hall.

Greenhouse, Steven. 2003. I.B.M. Explores Shift of White-Collar Jobs Overseas. New York Times: July 22, 2003.

Griliches, Z. 1990. Patent Statistics as Economic Indicators: A Survey. Journal of Economic Literature, 28(4): 1661-1707.

Hall, Bronwyn H., Adam B. Jaffe and Manuel Trajtenberg. 2001. The NBER Patent Citation Data File: Lessons, Insights and Methodological Tools. NBER Working Paper, No. 8498.

Hall, Bronwyn H., Adam B. Jaffe and Manuel Trajtenberg. 2003. Market Value and Patent Citations. Revised version of Market Value and Patent Citations: A First Look 1999. NBER Working Paper, No. 6984.

Hall, Bronwyn H. and Rosemarie Hm Ziedonis. 2001. The Patent Paradox Revisited: An Empirical Study of Patenting in the U.S.. The Rand Journal of Economics, 32(1): 101-128.

Hausman, Jerry, Bronwyn H. Hall and Zvi Griliches. 1984. Economic Models for Count Data with an Application to the Patent-R&D Relationship. Econometrica, 52(4): 909-938.

Henisz, Witold J. 2002. The Institutional Environment for Infrastructure Investment. Industrial and Corporate Change, 11(2): 355-389.

Jaffe, Adam B. 1986. Technological Opportunity and Spillovers of R&D: Evidence from Firms' Patents, Profits, and Market Value. American Economic Review, 76(5): 984-1001.

Jaffe, Adam B., Manuel Trajtenberg and Rebecca Henderson. 1993. Geographic localization of knowledge spillovers as evidenced by patent citations. Quarterly Journal of Economics, 434: 578-98.

Jaffe, Adam B., Manuel Trajtenberg and Michael S. Fogarty. 2000. The Meaning Of Patent Citations: Report On The NBER/Case-Western Reserve Survey Of Patentees. NBER Working Paper, No. 7631.

Kanellos, Michael. 2002. China: Bursting with Brainpower. ZDNet News: July 10, 2002.

Kaufmann, Daniel, Aart Kraay and Pablo Zoido-Lobatón. 1999. Governance Matters. The World Bank Development Research Group Working Papers, No. 2196.

Khanna, Tarun. 2000. Business Groups and Social Welfare in Emerging Markets: Existing Evidence and Unanswered Questions. European Economic Review, 44(4-6): 748-61.

Khanna, Tarun and Krishna G. Palepu. 1997. Why Focused Strategies May Be Wrong for Emerging Markets. Harvard Business Review, 75(4): 41-51.

Kogut, Bruce and Udo Zander. 1993. Knowledge of the Firm and the Evolutionary Theory of the Multinational Corporation. Journal of International Business Studies, 24(4): 625-45.

Kuemmerle, Walter. 1997. Building effective R&D capabilities abroad. Harvard Business Review, 75(2): 61–70.

Kuemmerle, Walter. 1999. The Drivers of Foreign Direct Investment into Research and Development: an Empirical Investigation. Journal of International Business Studies: 1-24.

Kumar, Nagesh. 2001. Determinants of location of overseas R&D activity of multinational enterprises: the case of U.S. and Japanese corporations. Research Policy, 30(1): 159-74.

Kynge, James. 2002. Rich Vein of Raw Talent Makes China Potential R&D Hothouse: Multinationals Can Cash in on Intellectual Benefits. Financial Times: April 19, 2002.

Lahiri, Nandini. 2003. Knowledge Spillovers: Geography, Technology and Intra-firm Linkages. Working Paper.

Levin, Richard C., Alvin K. Klevorick, Richard R. Nelson, and Sidney G. Winter, Richard Gilbert, Zvi Griliches. 1987. Appropriating the Returns from Industrial Research and Development. Brookings Papers on Economic Activity, 1987(3): 783-831.

Lee, Jeong-Yeon and Edwin Mansfield. 1996. Intellectual Property Protection and U.S. Foreign Direct Investment. The Review of Economics and Statistics, 78(2): 181-86.

Mansfield, Edwin, Mark Schwartz and Samuel Wagner. 1981. Imitation Costs and Patents: an Empirical Study. Economic Journal, 91: 907-18.

Morck, Randall and Bernard Yeung. 1991. “Why investors value multinationality,” Journal of Business, 64(2): 165-87.

Patel, Pari and K Pavitt. 1995. Patterns of technological activity: their measurement and interpretation. Stoneman, P. (Ed.), Handbook of the Economics of Innovation and Technological Changes. Blackwell, Oxford.

Patel, Pari and Modesto Vega. 1999. Patterns of internationalisation of corporate technology: Location vs. home country advantages. Research Policy, 28(2-3): 145-55.

Rapp, Richard T. and Richard P. Rozek. 1990. Benefits and Costs of Intellectual Property Protection in Developing Countries. Journal of World Trade, 75: 78-81.

Rivkin, Jan W. 2000. Imitation of Complex Strategies. Management Science, 46(6): 824-844.

Teece, David J. 1977. Technology Transfer by Multinational Firms: The Resource Cost of Transferring Technological Know-How. Economic Journal, 87(346): 242-261.

Teece, David J. 1986. Profiting from Technological Innovation: Implications for Integration, Collaboration, Licensing and Public Policy. Research Policy, 15(6): 285-305.

Vuong, Quang H. 1989. Likelihood Ratio Tests for Model Selection and Non-Nested Hypotheses. Econometrica, 57(2): 307-333.

Winter, Sidney. 2000. Appropriating the Gains from Innovation. G. Day and P. Schoemaker, eds. Wharton on Managing Emerging Technologies. New York: Wiley, 2000: 242-265.

TABLES

Table 1. Description of the Sample Patents

Year	Patents developed in weak IPR countries		Patents developed in all foreign countries		Total number of patents
	Number	Percentage	Number	Percentage	
1993	118	0.60%	2,070	10.53%	19,652
1994	209	0.98%	2,219	10.41%	21,313
1995	244	0.81%	3,259	13.50%	24,143
1996	277	0.95%	3.039	11.48%	26,474
1997	385	1.12%	3,406	9.91%	34,357
1998	386	1.22%	3,203	10.10%	31,708
1999	505	1.70%	3,332	11.21%	29,716
2000	437	1.79%	2,982	12.20%	24,434
2001	267	1.75%	1,820	11.94%	15,238
Total	**2,828**	**1.20%**	**25,330**	**11.16%**	**227,034**

Table 2. Description of the Sample Firms

Variables	Firms that do R&D		All firms (1567 obs)
	in weak IPR countries (227 obs)	in any foreign countries (681 obs)	
Assets (million dollars)	10,414.65	4,981.98	2,687.09
Sales (million dollars)	8,088.27	3,813.81	1,931.97
Number of patents[1]	750.14	298.60	137.61
Number of subsidiaries[1]	45.69	31.99	19.36
Number of assignees[1]	4.98	3.34	2.27
Countries with presence[1]	13.56	8.92	4.55
H-index for tech class[2]	0.08	0.12	0.28

[1] count per firm

[2] concentration of technology field, calculated as a Herfindahl index of the firm's patents across patent classes. Smaller numbers indicate more technology diversification.

Table 3. Self-Citation Ratios of Patents Developed under Different Environments

Year	Patents developed in weak IPR countries		Patents developed in other foreign countries		Patents developed in the home country	
	mean	s.d.	mean	s.d.	mean	s.d.
1993	0.124	0.183	0.123	0.250	0.166	0.257
1994	0.145	0.241	0.141	0.254	0.171	0.263
1995	0.170	0.257	0.120	0.247	0.175	0.274
1996	0.177	0.289	0.158	0.283	0.187	0.294
1997	0.175	0.295	0.153	0.288	0.206	0.318
1998	0.195	0.319	0.191	0.339	0.237	0.357
1999	0.242	0.385	0.241	0.389	0.279	0.392
2000	0.352	0.462	0.264	0.409	0.332	0.435
2001	0.556	0.498	0.458	0.502	0.391	0.463
Average	**0.191**	0.308	**0.161**	0.296	**0.205**	0.317

Table 4. Zero-inflated Negative Binomial Regression on Within-Firm Difference

	The Full Sample (1)	Firms with >50 patents (2)	Voluntary citations (3)	Control for patent class (4)	Computer & Telecom (5)
Foreign country	-0.2838*** (0.0476)	-0.2849*** (0.0477)	-0.2248*** (0.0397)	-0.2832*** (0.0424)	-0.4496*** (0.0503)
Weak IPR country	**0.2055*** (0.0703)**	**0.2033*** (0.0707)**	**0.2128*** (0.0865)**	**0.2207*** (0.0723)**	**0.3941*** (0.1059)**
Constant	-1.6743*** (0.2069)	-1.6755*** (0.2068)	-1.9263*** (0.1803)	-1.4925* (0.7644)	-1.6564** (0.9354)
Total citation	- exposure				
Inflate					
Total citation	0.0118*** (0.0029)	0.0118*** (0.0028)	0.0883*** (0.0107)	0.0137*** (0.0027)	0.0178*** (0.0052)
Constant	-2.4389*** (0.2678)	-2.4405*** (0.2678)	-2.9442*** (0.2682)	-2.4975*** (0.2630)	-3.2100*** (0.2711)
Obs	125,796	125,036	95,302	117,120	42,801
log_likelihood	-153,479.80	-153,106.10	-66,659.79	-147, 920.87	-52,536.13

*** significant at 1% level ** significant at 5% level * significant at 10% level

Numbers in parentheses ( ) are robust standard errors adjusted for clustering.

Table 5. Self-Citation Ratios of Patents Developed by Different Firms

Year	Firms with positive patent output in the 10-year period			Firms with >50 patents in the 10-year period	
	w/ R&D in weak IPR countries	w/ R&D in other foreign countries	w/o any foreign R&D	w/ R&D in weak IPR countries	w/o R&D in weak IPR countries
1993	0.175	0.129	0.083	0.180	0.129
1994	0.180	0.130	0.076	0.185	0.109
1995	0.163	0.117	0.070	0.167	0.134
1996	0.195	0.109	0.078	0.201	0.105
1997	0.219	0.140	0.101	0.226	0.164
1998	0.249	0.179	0.166	0.254	0.185
1999	0.294	0.216	0.212	0.299	0.281
2000	0.340	0.251	0.365	0.347	0.273
2001	0.410	0.315	0.396	0.408	0.348
Average	**0.211**	**0.144**	**0.125**	**0.216**	**0.153**

Table 6. Zero-inflated Negative Binomial Regression on Cross-Firm Difference

	The Full Sample (1)	Firms with >50 patents (2)	Voluntary Citations (3)	Control for patent class (4)	Computer & Telecom (5)
Firms w/ R&D in foreign	0.3196** (0.1322)	-0.0963 (0.1946)	0.1016 (0.1677)	0.3112** (0.1212)	0.7188*** (0.2262)
Firm w/ R&D in weak IPR	**0.3950*** (0.1122)**	**0.3453*** (0.1168)**	**0.5337*** (0.1169)**	**0.3239*** (0.0812)**	**0.5256*** (0.1427)**
Developed in foreign	-0.3811*** (0.1526)	-0.3743** (0.1560)	-0.2753* (0.1453)	-0.4605*** (0.1118)	-0.6570** (0.2195)
Developed in weak IPR	0.0390 (0.1982)	0.0578 (0.2020)	0.0573 (0.2166)	0.2002 (0.1307)	0.4846 (0.3003)
Constant	-2.2573*** (0.1187)	-1.7852*** (0.1842)	-2.8044*** (0.3114)	-2.5436*** (0.5478)	-3.0136*** (0.2133)
Total citation	- exposure				
inflate					
Total citation	0.0102*** (0.0037)	0.0095*** (0.0039)	0.0830* (0.0045)	0.0097*** (0.0039)	0.0145*** (0.0032)
Constant	-2.0344*** (0.1977)	-2.0438*** (0.2057)	-1.9950 (1.3005)	-1.9433*** (0.2363)	-2.7041*** (0.2400)
Obs	153,950	146,018	116,138	153,950	49,733
log_likelihood	-191,527.2	-185,673.7	-83,201.8	-186,562.4	-62,134.3

*** significant at 1% level ** significant at 5% level * significant at 10% level

Numbers in parentheses ( ) are robust standard errors adjusted for clustering.

Table 7. Cross-Firm Regression with Size Effect

	Control for Assets (1)	Control for Sales (2)	Control for Patent Output (3)
Assets ($bn)	-0.0020*** (0.0010)		
Sales ($bn)		-0.0024* (0.0018)	
Patent Output (thousand)			0.0170* (0.0093)
Firms w/ R&D in weak IPR	**0.4761*** (0.1364)**	**0.4788*** (0.1474)**	**0.3018** (0.1343)**
Firms w/ R&D in foreign	0.3415*** (0.1657)	0.3422*** (0.1656)	0.2625** (0.1641)
Developed in weak IPR	0.0340 (0.1664)	0.0410 (0.1480)	0.0783 (0.2070)
Developed in foreign	-0.3655*** (0.1763)	-0.3686** (0.1745)	-0.4048** (0.1761)
Constant	-2.2574*** (0.1506)	-2.2638*** (0.1512)	-2.1047*** (0.1429)
Total citation	- exposure		
inflate			
Total citation	0.0089** (0.0047)	0.0088** (0.0048)	0.0138*** (0.0032)
Constant	-2.1412*** (0.3043)	-2.1321*** (0.3181)	-2.2353*** (0.2764)
Obs	126,636	126,636	153,950
log_likelihood	-160,355.50	-160,369.40	-192,347.00

*** significant at 1% level ** significant at 5% level * significant at 10% level

Numbers in parentheses ( ) are robust standard errors adjusted for clustering.

Table 8. Within- and Cross-Firm Regression with Stable Firm Affiliations

	Within-firm (1)		Across-firm (2)	
Firms w/ R&D in weak IPR			**0.4309*** **	**(0.1356)**
Firms w/ R&D in foreign			0.3115**	(0.1776)
Developed in weak IPR	**0.2543*** **	**(0.0831)**	0.0826	(0.2324)
Developed in foreign	-0.3034***	(0.1019)	-0.4131**	(0.1830)
Constant	-1.6456***	(0.1933)	-2.2466***	(0.1548)
Total citation	- exposure			
inflate				
Total citation	0.0105***	(0.0039)	0.0095***	(0.0038)
Constant	-2.5555***	(0.1899)	-2.1863***	(0.2452)
Obs	106,365		129,614	
log_likelihood	-130,730.00		-162,695.30	

*** significant at 1% level ** significant at 5% level * significant at 10% level

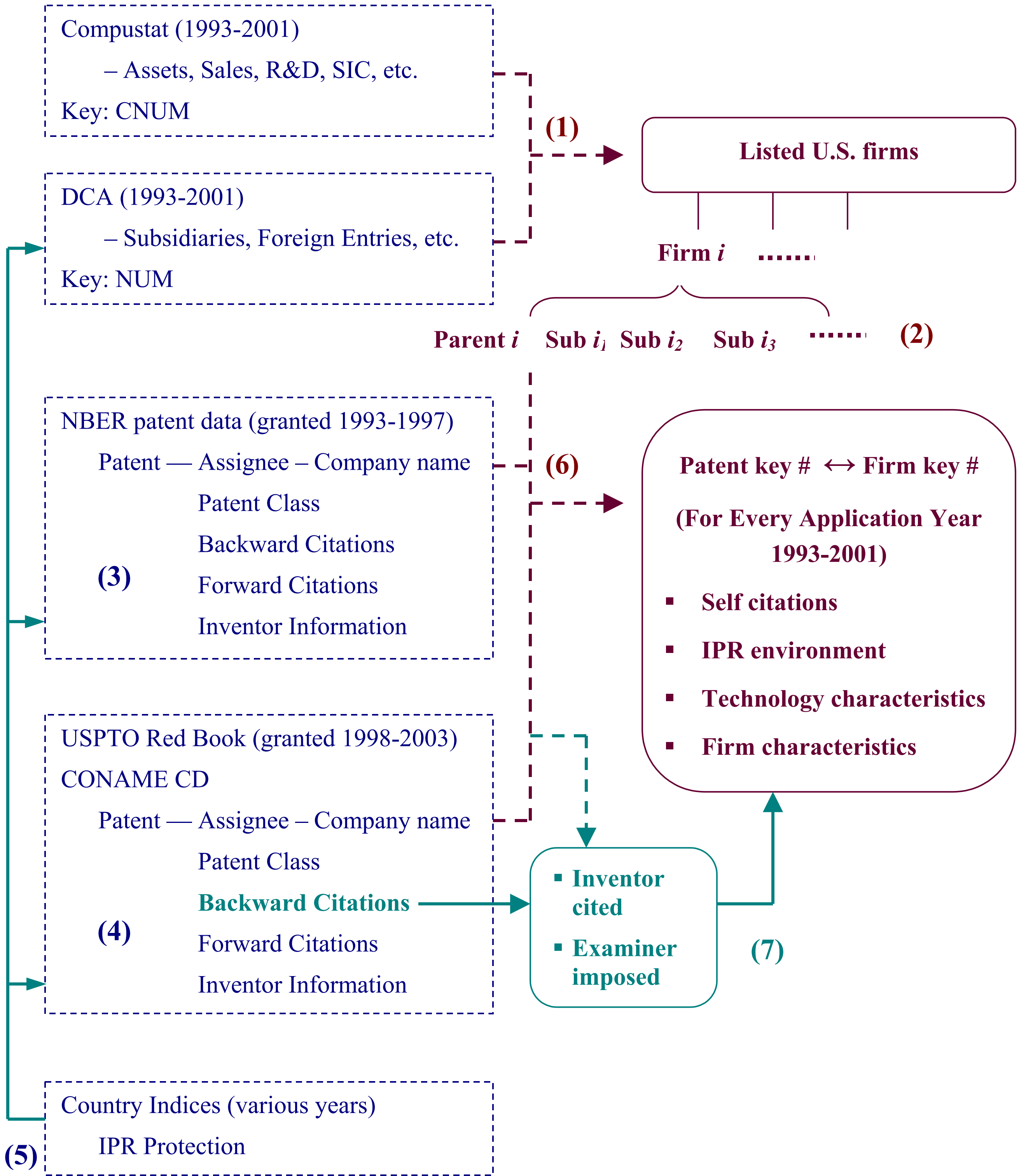
Numbers in parentheses ( ) are robust standard errors adjusted for clustering.

APPENDIX A. Countries with Weak IPR Protections

	Scientist &Engineers /mn people	Population (mn)	Tertiary School (gross %)	Opacity Factor 2000	Property Rights 1995	Law & Order 93-95	Rapp & Rozek	Ginarte & Park 1990	Special 301* 1999	KKZ Index 1998	Piracy Rate 2000
Argentina	711	35.22	41.80	60.60	2	4.58	1	2.26	1.5	0.24	58
Belarus	2,296	10.30	44.00		3				1.0	-1.08	
Brazil	168	161.52	11.70	60.85	3	3.25	1	1.85	1.0	-0.09	58
Bulgaria	1,289	8.36	41.20		3					-0.22	
Chile	370	14.42	30.30	35.65	1	4.58	2	2.41	1.0	1.26	49
China	459	1,215.30	5.70	87.16	4		1		2.0	-0.22	94
Costa Rica	533	3.40	33.10		3		3	1.47	1.0	0.88	
Czech Rep.	1,317	10.32	22.70	70.81	2				1.0	0.62	43
Egypt	493	59.27	22.60	57.97	4	3.61	2	1.99	1.5	0.17	
Greece	1,045	10.48	42.80	57.38	2	5.56	4	2.32	1.5	0.66	66
HKSAR, China	93	6.32	28.00	44.68	1	5.36		2.57		1.73	57
Hungary	1,249	10.19	25.10	50.07	2				1.0	0.78	51
India	158	945.78	6.90	63.74	3	3.83	1	1.48	1.5	0.21	63
Indonesia	..	197.18	11.30	75.16	3	4.22	0	0.33	1.5	-0.97	89
Israel	1,570	5.69	43.60	52.71	2	5.00	5	3.57	1.5	1.09	41
Korea, Rep.	2,139	45.51	60.30	73.46	1	5.00	3	3.94	1.0	0.82	56
Lithuania	2,031	3.71	31.40	58.45						0.19	
Malaysia	154	21.14	11.40		2	4.61	3	2.37		0.82	66
Mexico	213	92.71	16.10	47.64	2	3.00		1.63	1.0	-0.38	56
Pakistan	78	125.42	3.40	61.96	2	2.64		1.99	1.0	-0.71	
Peru	229	23.95	31.10	57.63	3	2.83	1	1.02	1.5	-0.44	61
Philippines	156	71.90	35.20		3	3.78	4	2.67	1.0	-0.04	61
Poland	1,460	38.62	24.30	63.93	3				1.0	0.57	54
Portugal	1,583	9.93	38.00		2	5.42	3	1.98		1.31	42
Romania	1,393	22.61	22.50	71.42	4				1.0	-0.25	
Russia	3,397	147.73	41.40	83.59	3				1.5	-0.78	88
Slovak Rep.	1,706	5.35	22.10		2					0.13	
South Africa	992	39.90	18.80	59.54	3	3.33	5	3.57	1.0	0.21	45
Spain	1,562	39.27	51.10		2	6.00	4	2.95	1.0	1.35	51
Taiwan, China	660	21.42	18.71	60.64	1	5.00			1.0	1.17	53
Thailand	102	60.00	20.90	66.95	1	5.00	1	1.85	1.0	0.40	79
Turkey	303	61.45	18.20	74.07	2	4.17	1	1.80	1.5	0.19	63
Ukraine	2,121	51.11	41.50		4				1.5	-0.76	
Venezuela	194	22.31	25.40	63.45	3	4.00	2	1.35	1.0	-0.62	58
United States	4,103	265.23	80.60	35.53	1	6.00	.	4.52	.	1.77	26

* The countries with a "1" are countries on the watch list, those with a "1.5" are on the priority watch list, and those with a "2.0" are section 306 monitoring countries.

APPENDIX B. Illustration of Data Sources



Notes:

- Dotted rectangular: data sources
- Rounded rectangular: processed datasets
- Dashed lines: data matching
- Solid lines: information reference

Steps

(1) Match Compustat with DCA data, year by year, according to company names.

(2) For each matched firm, extract all the family members from DCA.

(3) Prepare data for patents applied on or after 1993, and granted between 1993 and 1997, using the NBER dataset.

(4) Prepare data for patents applied on or after 1993, and granted between 1998 and August 2003, using the USPTO data. Assignee names are modified according to the USPTO company name files.

(5) Country indices are used to describe the countries of the inventors.

(6) Patent assignee names are matched with all the company names –parents as well as subsidiaries and other family members –in the Compustat-DCA company list. Thus, every patent (with relevant information) is corresponding to a firm (with relevant information). Inventor locations are used to determine the IPR environment in which the technology is developed. Self-citations are used to proxy for internal linkages.

(7) Among the self-citations from (6), count the number of self-citations imposed by the examiner. Exclude those citations in the robustness check.