



University of Pennsylvania
ScholarlyCommons

Management Papers

Wharton Faculty Research

8-2006

Conducting R&D in Countries With Weak Intellectual Property Rights Protection

Minyuan Zhao
University of Pennsylvania

Follow this and additional works at: https://repository.upenn.edu/mgmt_papers

 Part of the [Management Sciences and Quantitative Methods Commons](#)

Recommended Citation

Zhao, M. (2006). Conducting R&D in Countries With Weak Intellectual Property Rights Protection. *Management Science*, 52 (8), 1185-1199. <http://dx.doi.org/10.1287/mnsc.1060.0516>

This paper is posted at ScholarlyCommons. https://repository.upenn.edu/mgmt_papers/298
For more information, please contact repository@pobox.upenn.edu.

Conducting R&D in Countries With Weak Intellectual Property Rights Protection

Abstract

Multinational enterprises (MNEs) are increasingly conducting research and development (R&D) in countries such as China and India, where intellectual property rights (IPR) protection is still far from adequate. This paper examines this seemingly puzzling situation. I argue that weak IPR leads to low returns to innovation and underutilization of innovative talents; MNEs that possess alternative mechanisms for protecting their intellectual properties will therefore find it attractive to conduct R&D at those locations. A theoretical framework is developed to capture the interaction between firm strategy and the institutional environment. The empirical analysis on a sample of 1,567 U.S.-headquartered innovating firms finds results consistent with the hypotheses that (i) technologies developed in countries with weak IPR protection are used more internally, and (ii) technologies developed by firms with R&D in weak IPR countries show stronger internal linkages. The results suggest that firms may use internal organizations to substitute for 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, MNEs, arbitrage

Disciplines

Management Sciences and Quantitative Methods

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

Minyuan Zhao*

Stern School of Business

New York University

mzhao@stern.nyu.edu

September 2003

Preliminary

* Graduate student, Stern School of Business, New York University. 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 David Backus, Thomas Cooley, seminar participants at Carnegie Mellon University, New York University, CCC, AIB Doctoral Consortium, and BPS Doctoral Consortium for helpful comments and suggestions. All errors remain mine. I appreciate the support from 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

Multinationals are increasingly conducting R&D in countries such as India and China, where intellectual property rights (IPR) protection is still very weak. This paper examines this puzzle. The argument is that weak IPR leads to low returns to innovation and thus low prices of innovative talents. Multinational firms who possess not only the capabilities to utilize these talents, but also the internal organizational structures 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 stylized model of the above is presented to capture the interaction between firm strategies and institutional environment. Empirical findings from a sample of 1567 US-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 are using strong internal linkages to substitute for the inadequate external institutions. And by doing so, they can take advantage of the arbitrage opportunities presented by the institutional gaps across countries.

Keywords: R&D, intellectual property rights, multinational firms, 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 General Electric are in the lead, but more firms are following. (Financial Times 4/19/02; New York Times 4/21/02; ZDNet News 6/10/02; BusinessWeek 2/03/03; Wall Street Journal 7/14/03) 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 in the multinational R&D labs in China. Some common practices emerged: intensive interactions with headquarters, patents application in the home country, and internal project transfers across countries. In particular, the projects are often closely integrated in the multinational 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 cannot appropriate value from their 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 talents to conduct R&D. Multinationals 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 it the *internalization-arbitrage* conjecture.

To articulate the idea in a logically traceable way, a stylized model of the above is presented. It shows that multinationals may find it desirable to conduct R&D across borders when technologies are complementary internally. By keeping the complementary resources well protected, multinationals 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.

The study then seeks empirical evidence of the theoretical conjecture, using US 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 staying out. The results are consistent with the thought that the internal linkages allow firms to appropriate value from their knowledge even under weak institutional environment.

The following section briefly describes the interviews in China. Section III presents a simple model to show how arbitrage opportunities can emerge from the interaction between firms and institutions. Section IV brings the theoretical conjecture to the data and sets up the framework 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 size 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, multinationals are streaming into China to tap the best talents in the country. By the end of 2002, over four hundred foreign owned R&D centers have been established, hiring away the best masters and Ph.D.s from China's top universities and research institutes. What enable the multinationals to do something that the locals are not able to do? With this puzzle on mind, I conducted a series of interviews with researchers and managers in some major 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 the human capital and appropriate value from the R&D projects in China, multinationals have to keep the knowledge leakage at the lowest level possible. At the same time, China has a booming domestic market. Competition for market shares requires that multinationals effectively transfer and adapt technologies for their local operations. This dilemma leads to a common practice among multinational firms 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". The mission of IBM China Research Lab (CRL) is to "create world-class information technologies and the underlying science which propel the world advances". Intel China

¹ Seven of the labs belong to Global 500 companies, and six have more than 100 researchers employed. Information has been updated through emails and phone calls during the year.

Software Lab (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 multinational'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 of these research labs facilitates the transfer of research projects across locations, and thus makes it possible to exercise the strategy of differentiated project assignment. 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, people in the research labs do not seem to be very concerned about imitation risks. When I asked the question: “Given the weak IPR protection in China, are you concerned that your technology 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 the 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 later forms the basis of the newest MSN Messenger package, is a frontier technology building on Microsoft's strength in computer vision and computer graphics. In the Intel lab, most efforts are on developing new BIOS, compilers, and device drivers for 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. But very few final products are developed in these Beijing or Shanghai labs. The research results will be integrated into the final applications somewhere else, most likely 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 BOIS, 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 is later integrated in “Microsoft Movie Maker”, which is distributed with the new Windows operating systems. Similar examples are the “Mobile HTML Optimizer” used in Microsoft FrontPage, the “Ink Parsing” technology used in Tablet PC, and the “error-resilient video transmission” technology used in the MPEG4 Standard. 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 told me.

In sum, if technologies inside the firm are highly complementary, the leakage of a particular technology will 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 the integration with 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 some other weak IPR countries — has just opened its door to foreign investors. Even

after the opening up, for a long time the government required minimum local stakes in foreign invested enterprises, which made close integration almost impossible. In fact, nearly all these research labs were established after restrictions on wholly owned subsidiaries were removed.

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

Multinational firms also gained more experience in organizing large-scale R&D projects and in dealing with the institutional idiosyncrasy in China. “Multinationals are gradually learning how to move smartly, and some learned their lessons in the hard way,” a researcher said when referring to the firm’s earlier loss from counterfeits.

III. THE MODEL

The interviews suggest that firms may be able to use their internal organizations to protect knowledge, hence taking advantage of the inexpensive human capital in weak IPR countries. In this section I present a simple model to show how technology complementarities, firm organizations, and legal institutions interact with one another. This model serves two purposes: to help me study the cross-border arbitrage in a more 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 the 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 flows.

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 only happen 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 who lack the resources to utilize the knowledge.

Second, the acquisition of complementary knowledge is subject to the constraints of absorptive capacity and 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 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 multinationals 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 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 two-country two-projects model.

3.2 The Model Setup

To focus on the *organization* of R&D activities, the model assumes away other factors that may affect the generation and appropriation of intellectual properties. For instance, large multinational firms, with their worldwide production and marketing networks, are

in a better position to use 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 less financial constraint. These factors are left out of the model. Also for simplicity, I assume Bertrand competition among identical products, although the framework applies to more general settings.

Suppose that firm A has two complementary technologies under management: a_1 and a_2 . Each technology a_i ($i = 1, 2$) has stand-alone value $V(a_i)$. For example, Sharp Corporation's recent innovation, the Continuous Grain Silicon (CG-Silicon) technology, provides higher resolution and brightness to compact LCDs. Hence, this technology by itself has value $V(a_i)$, which can be considered the rent that Sharp is able to collect from LCD manufacturers. Once imitation occurs, Sharp would lose its pricing power on this technology and $V(a_i)$ would plummet to zero.

At the same time, a_1 can also be integrated with a_2 as complementary technologies 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 $V(a_1 + a_2)$, or $V(A)$, is different from the simple sum of individuals $V(a_1) + V(a_2)$. Let

$$\delta = \frac{V(a_1) + V(a_2)}{V(a_1 + a_2)}$$

I argue that δ is less than one for two reasons: First, technologies developed in one firm may not be readily applicable to other firms who possess a different set of resources. So the stand-alone value $V(a_i)$ can 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 generally describe the difference between the joint value $V(a_1 + a_2)$ and the external value $V(a_1) + V(a_2)$.

Note that this is a very general framework. Internal complementarity – hence higher internalized value – 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. Let $\delta_1 = V(a_1) / V(A)$ and $\delta_2 = V(a_2) / V(A)$, then $\delta_1 + \delta_2 = \delta < 1$. Note that even within the same firm, the degree of internal dependence may vary across different types of knowledge ($\delta_1 \neq \delta_2$).

Depending on the legal and social institutions, there is certain imitation risk p ($0 \leq p \leq 1$) in the economy, which is taken as given by individual firms. With probability p , imitation would happen to technology a_i and the stand-alone 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 $V(A)$.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 many 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. With the proprietary 3GT polymer that a_2 is designed for, a_2 is much more valuable to Dupont than to any other firms.

² Even if both technologies are imitated, it is still uncertain whether the imitator can perfectly replicate the integration process. The development of these technologies involves intensive interaction with each other and with the firm's existing expertise, so the synergy is not readily transferable. This factor is abstracted from the model to make the main theme more traceable. Further analysis shows that although partial replication can be interesting of its own, omitting it from the model does not change the results.

On the cost side, each project incurs R&D expense $C(a_i)$. Think of $C(a_i)$ as the payment to scientists and engineers who charge market price c for their human capital. The prevailing market price c is determined by the expected returns from innovation in the economy: the higher the imitation risk p , the smaller is the value appropriable by the innovators, and thus the lower is the price of human capital. For individual firms, c is taken as exogenous. Firms will invest in R&D if and only if the appropriable value is higher than the costs.

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 boundary, complementary technologies create synergy in a way that is hard to duplicate outside the firm. Within a national boundary, R&D activities are subject to certain institutional environment that is distinct from that of other countries. Multinationals are of particular interest because they expand their firm 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 . Firm A can choose where to develop its two complementary projects. It can follow one of 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 in country Y .

As specified above, the cost of developing technology a_i in country j ($j = x$ or y) is the amount of human capital used (H_i), times the market price for human capital (c_j).

³ Even within one country, firms can choose different organizational structures for their R&D, which in turn depend on the external environment. For example, R&D by small startups is common in the US, 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 will be studied in depth in my future work.

In addition, cross border R&D incurs extra coordination and communications costs R (Kuemmerle, 1997), which is decreasing in the firm’s organizational capability. Let $R = rV(A)$, where r reflects the organizational cost of the firm, or organizational inefficiency. Since $V(A)$ is the value – not only the size – of the projects, an implicit assumption here is that higher synergy would also require higher coordination costs.

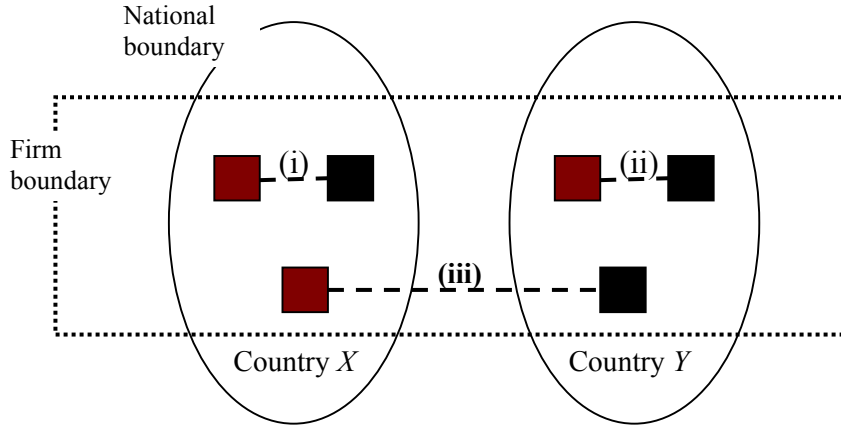


Figure 1. Organizational Strategies of R&D Projects

The expected income from different R&D strategies depends on the value creation as well as the ability to appropriate the created value. As an example, the following table shows the risk profile of strategy (iii), in which a_1 is developed in country X and a_2 is developed in country Y :

Probability	Value loss	Comments
$(1 - p_x) (1 - p_y)$	0	No imitation
$p_x (1 - p_y)$	$V(a_1)$	Losing the stand-alone value of a_1
$(1 - p_x) p_y$	$V(a_2)$	Losing the stand-alone value of a_2
$p_x p_y$	$V(A)$	Losing the joint value

So the expected appropriable value is:

$$V = V(A) - p_x V(a_1) - p_y V(a_2) - p_x p_y (1 - \delta) V(A)$$

Intuitively, V is the created value $V(A)$ minus the expected loss of stand-alone values and — if both technologies are imitated — the internalized value. Throughout this model I

assume risk neutral decision-making, which is purely for the purpose of simplification. Taking risk averse into consideration will not qualitatively change the analysis.

Combining V with the total cost

$$C = C(a_1) + C(a_2) + R$$

we get the expected profits

$$\pi = V - C$$

The following table presents the expected profits of all three strategies, equally scaled by the joint value $V(A)$.

#	Strategy	Return
(i)	Both in country X	$\pi^{(i)} = 1 - p_x \delta - p_x^2 (1 - \delta) - (h_1 + h_2) c_x$
(ii)	Both in country Y	$\pi^{(ii)} = 1 - p_y \delta - p_y^2 (1 - \delta) - (h_1 + h_2) c_y$
(iii)	Each in one country	$\pi^{(iii)} = 1 - p_x \delta_1 - p_y \delta_2 - p_x p_y (1 - \delta) - h_1 c_x - h_2 c_y - r$

Here $h_i = H_i / V(A)$: the amount of human capital for each unit of value created. Because human capital can be measured in arbitrary units, I normalize $h_1 + h_2$ to 1 in the following analysis.

3.4 Internalization-Arbitrage

Now we are ready to analyze the potential arbitrage opportunities in strategy (iii), compared with purely domestic plays (i) and (ii). In strategy (iii), the two complementary components are developed in two different environments with a institutional gap $\Delta p \equiv p_y - p_x > 0$. The model will show that, given the difference in factor prices $\Delta c \equiv c_x - c_y > 0$,⁴ firms with the right internal capabilities are able to gain from the cross-border R&D arrangement.

First, why would firm A in country X have the incentive to go abroad and develop a project in country Y ? Take the difference between (iii) and (i):

⁴ Theoretically, the gap in factor prices will be endogenously affected by firm's cost arbitrage activities. But 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 will be analyzed in another chapter of my dissertation.

$$\pi^{(iii)} - \pi^{(i)} = -\underbrace{[\delta_2 + p_x(1-\delta)] \cdot \Delta p}_{(1)} + \underbrace{h_2 \cdot \Delta c}_{(2)} - \underbrace{r}_{(3)} \quad (a)$$

In this equation, part (1) is the extra loss due to the additional imitation risk Δp . It consists of the potential loss of stand-alone value δ_2 as well as the loss of internalized value $(1-\delta)$. What is intriguing here is that, given Δp , the loss of internalized value is reduced if the home country has very strong IPR protection (small p_x). The reason is that $(1-\delta)$ stays within the firm as long as *either* of the two technologies remains exclusive. Hence, a strong home institution is critical to the protection of the internalized value. The interaction between firms and institutions comes into play in the presence of strong technology complementarity within multinational firms.

Compensating for the extra risks, part (2) is the gain from the inexpensive human capital in country Y . The cost savings are large if the project developed there needs a large amount of human capital. Of course, cross-border R&D arrangement incurs the extra coordination cost r , which has to be justified by the advantage derived from (2) – (1).

Second, why wouldn't firm A conduct all the R&D in the low cost country? Take the difference between (iii) and (ii):

$$\pi^{(iii)} - \pi^{(ii)} = \underbrace{[\delta_1 + p_y(1-\delta)] \cdot \Delta p}_{(1)} - \underbrace{h_1 \cdot \Delta c}_{(2)} - \underbrace{r}_{(3)} \quad (b)$$

The intuitions behind (b) are similar to those behind (a). Part (1) is the gain from strong IPR protection in the home country. Since a_1 is developed in country X , the stand-alone value δ_1 is better protected and the internalized value $(1-\delta)$ is also subject to smaller imitation risks. The worse the IPR protection in the host country (large p_y), the more advantageous it is to be multinational. 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.

However, having a_1 developed in country X also means higher cost of human capital. Part (2) reflects the cost disadvantage compared with purely domestic firms in country Y . The extra coordination cost r may also undo the gains from (1) – (2), and therefore discourage cross-border R&D.

Equations (a) and (b) explicitly point out the potential arbitrage opportunities in multinational R&D activities. The cross-border arrangement would be a desirable choice if $\pi^{(iii)} > \pi^{(ii)}$ and $\pi^{(iii)} > \pi^{(i)}$ hold simultaneously, i.e.,

$$\begin{cases} \pi^{(iii)} - \pi^{(i)} = -[\delta_2 + p_x(1-\delta)] \cdot \Delta p + h_2 \cdot \Delta c - r > 0 \\ \pi^{(iii)} - \pi^{(ii)} = [\delta_1 + p_y(1-\delta)] \cdot \Delta p - h_1 \cdot \Delta c - r > 0 \end{cases}$$

Rearrange the terms:

$$\begin{cases} (-\delta_2 \cdot \Delta p + h_2 \cdot \Delta c) - p_x(1-\delta) \cdot \Delta p - r > 0 & \text{(a')} \\ (\delta_1 \cdot \Delta p - h_1 \cdot \Delta c) + p_y(1-\delta) \cdot \Delta p - r > 0 & \text{(b')} \end{cases}$$

Given Δp and Δc , (a') and (b') can hold simultaneously if:

- (1) The project developed in country Y takes a large amount of human capital h_2 so that the cost savings are large, but it has limited stand-alone value δ_2 so that the potential risks are small. Meanwhile, the project developed in country X has a high stand-alone value δ_1 , but it requires relatively small amount of human capital h_1 .
- (2) The imitation risk p_x is low so that having one project developed in country Y brings little damage to the internalized value $(1-\delta)$. Meanwhile, the imitation risk p_y is high so that keeping one project in country X is critical to the protection of $(1-\delta)$.
- (3) The organizational cost r is reasonably low.

The three components are not mutually independent. For example, the organizational cost affects the R&D structure endogenously chosen by the firm, and the feature of individual components affect the overall degree of internalization. However, the model provides me a logically traceable framework to examine the original puzzle. Essentially, firms can strategically use their *internal organizations* to *arbitrage* the institutional gap across countries, which I call the *internalization-arbitrage* mechanism.

3.5 Insights from the Model

We were faced with a puzzle: why do multinationals 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 realize value from their innovative activities. Human capital is underpriced as a result. Such low cost human capital is attractive to R&D intensive firms if they possess alternative mechanisms to protect their intellectual property. One of the mechanisms, as identified above, is the internal complementarity of technologies.

First, R&D projects that are heavily reliant on internal complementarities are less appealing to imitators. This enables firms to conduct R&D in weak IPR countries without exposing themselves to too much risk. For example, software module 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 multinationals' internal R&D architectures. In contrast, projects developed in the home country, such as system designs or core products, tend to be more valuable on their own. The strategy of *differentiated project assignment*, in essence, is to tailor the projects so that the firm can capitalize on the strengths of particular locations while minimizing the costs or risks.

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. Usually in the IT industry, R&D projects are easier to decompose than in 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, multinationals, by keeping the key components under a strong IPR regime, are in a unique position to protect the *internalized value* arising from internal complementarities. The tighter their internal knowledge structures are, the more advantage they have compared with purely domestic firms in the host country.

The multilingual technology on computer systems is a good example of the internalized value. Because of the non-alphabetical nature of Chinese characters, Chinese software companies have been developing multilingual editors and related products since the early 1980s. They failed one after another due to prevailing piracies. 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, the value appropriation is not reliant on the small stand-alone value of specific components. Most local firms realize the value of multilingual technologies and are able to integrate the technologies in 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 the strong IPR countries. While local firms have to face prevailing piracy and slim profit for their products, multinationals can realize the value of these innovations in 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 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 can differ widely across firms, depending on the firms' previous exposure to institutional idiosyncrasies, their 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 ideas to the data and construct empirical measures to capture the theoretical concepts.

4.1 Implications of Internalization-Arbitrage

The key insight that emerges from the qualitative analysis is the substitution of firm's internal capabilities for the poor external environment. A multinational firm can take advantage of the inexpensive human capital in weak IPR countries if the value of the technologies can only be appropriated inside the firm, and if the firm can efficiently coordinate cross-border R&D activities.

Let γ_k be the measure of internal linkages for technology k , which is owned by firm i and developed in country j . According to the internalization-arbitrage conjecture, γ_k should depend on (1) the external IPR environment under which k is developed, (2) the characteristics of the firm that develops the technology. Within each firm, γ_k is expected to be higher if k is developed in a country with weak IPR protection. Across firms, firms that are able to conduct R&D in weak IPR countries are expected to have a higher γ_k for all the technologies they develop, everything else being equal. This lends a natural framework to the empirical analysis.

First, I focus on firms that conduct R&D in countries with weak IPR protection 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 firms and examine firm *capabilities*: who are doing the arbitrage? Internalization-arbitrage is a viable strategy only for firms with the right internal capabilities to defy the adverse environment in the host countries. 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 feature tighter organization of their innovative activities and more efficient coordination across countries.

These tests entail the following tasks:

- Define “weak IPR countries” using a series of widely cited indices;
- Describe the data sources and the sample;
- Construct variables from sample data to capture the theoretical concepts;
- Identify the right econometric model for the empirical analysis.

I will address the four tasks in detail in the next four subsections.

4.2 Defining Weak IPR Countries

Arbitrage opportunities exist when a country has a substantial reserve of human capital, yet suffers from weak IPR protection. Referring to the *World Development Indicators*, I remove countries and areas that have less than two million population, less than 1% of gross tertiary school enrollment rate, or less than five patents filed with USPTO in the whole 1990s. The removed countries and areas are, for example, Belize, Trinidad, and most sub-Sahara African countries. The reason for the screening is that the small size of human capital reserves will not justify the entry cost of multinationals. Although a handful of inventors do reside in these countries, they are very likely to represent extreme cases. Including these data may cause biases rather than adding explanatory powers. For the same reason, I remove war-torn countries such as Croatia, Cuba, 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. But given the slow changes in institutions, these indices should be indicative of the IPR protection levels in these countries.

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

- The Law and Order index in the *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 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 United States Chamber of Commerce. It covers about 97 countries, and pertains to the situation in the mid-80s. Out-of-date as it is, researchers and businesses are still using it as a reference to gauge the IPR environment of the host countries.
- Ginarte and Park (1997) index: This frequently cited index was produced for five-year intervals starting with 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.” Although these lists may be biased for political reasons, it is relevant to this study because I am studying the behaviors of American firms, who would probably refer to the government announcements before making their investment decisions.

I also supplement the indices with the Political Constraints measure developed by the Rule of Law index developed by Kaufmann, Kraay and Zoido-Lobaton (1999, 2002), and the piracy index developed by the International Planning and Research Corporation (IRC) on behalf of the Business Software Alliance (BSA) and the Software & Information Industrial 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 very stable list of 34 countries with weak IPR protection. The country names and the corresponding indices are listed in Table 1A in the Appendix. Interestingly, this list is not restricted to low-income regions. The per capita Gross National Income ranges from US\$25,920 for Hong Kong (China) to US\$440 for Pakistan in 2000, according to the World Bank statistics.

4.3 Data

Despite the various criticisms on patent data, I choose to use the US 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. Multinationals would only file US patents for frontier technologies that have potential value in the home country. With patent data instead of R&D expenditures, I can 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 kind of R&D as defined in Kuemmerle (1999). Moreover, because patents are the output of R&D, I can capture the projects that *fruitfully* utilize human capital in the weak IPR countries.

Second, patent citation is so far the most traceable evidence of knowledge flows (Jaffe, Trajtenberg and Fogarty 2000). The systematic documentation of patent citations tracks the knowledge flows within and across the firms' global innovation networks. From forward patent citations, I can identify who are following up the innovations, when, and where.

Finally, the detailed location information for all the patent inventors can help me identify the geographic distribution of talents utilized by American firms. Since the inventors' mailing addresses (not permanent addresses) are required in the patent application, they should reliably reflect the actual locations of the innovations.

Information on patents granted between January 1993 and December 1999 is obtained from the NBER patent data. Patents granted between 2000 and August 2003 are analyzed using the weekly Grant Red Book V2.5 bibliographic data from USPTO. Each entry field is closely examined to ensure consistency when the two data sets are merged.

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 by the applicants themselves. It would be unlikely that those imposed citations represent learning. I will analyze these problems in depth in the robustness check.

Since patents may be assigned to subsidiaries or parent companies 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. By an American *firm* I mean an ultimate parent registered in the United States, plus all its subsidiaries and affiliates, home or abroad, that are owned by the same ultimate parent, directly or indirectly, by more than 10%.⁵ After aggregation, the number of firms in my sample is only half of the number of assignees.

Compustat data are used to capture other firm level information, such as industry classification, size, etc. To avoid noises from small or instable firms, I drop companies with less than \$0.1 million assets, as well as those that are active for less than three years.

⁵ This is the official criterion of FDI for US companies. I will vary this number in the robustness check and make sure that the ownership threshold does not affect the result.

It is true that using Compustat data limits the study to the publicly traded companies, but it also eliminates the incomparability issue between public and private firms, as these two types of firms are subject to very different operational constraints. Given the fact public firms hold the majority of US patents (Hall, Jaffe, and Trajtenberg 2001), they should be representative of the innovating firms.

4.4 The Sample

I decide to focus on US-headquartered firms in the empirical study. The reason is that I can obtain the most complete data for US 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 that are applied during this period (granted up to August 2003). The reason why I do not include more recent years is that there is a typical 2-year time lag between patent application and patent granting. Using the most recently applied patents will bias the sample toward the selection of “quick patents”. At the same time, R&D in weak IPR countries is still a very recent phenomenon. Extending the sample period further back will not add much value to the analysis either. As robustness check, I also examine this sub-period from 1993 to 1997 in order to stay away from the irrational expansion during the Internet Bubble.

Different industries vary widely in their propensity to patent and the usefulness of patents as the indication of innovative activities (Cohen, Nelson and Walsh 2000). The industry control variables can only partly alleviate the problem. To reduce unrelated noises, I eliminate the industries where patents are very weak indicator of innovations until very recent years (e.g., insurance), industries that are heavily influenced by public policies (e.g., utility), industries that are domestic by nature (e.g., retailing), or industries whose geographic locations are dictated by some exogenous effects (e.g., mining). The main sample includes the following 2-digit SIC industries: 28 (chemical and allied products), 29 (petroleum and coal products), 30-39 (manufacturing), 48 (communications), 73 (business services, including software), and 87 (engineering and management services). I

shall vary the industry selection in the robustness check and make sure that our findings do not depend on specific industries.

The challenge in the data preparation is to match the major datasets together, where the only link among them is the company names. A computer program⁶ is developed to match the records together by company names. I decide not to use the Compustat match offered in the NBER patent data, which uses 1989 Compustat data. Instead, matching is conducted year by year to accommodate possible name changes — which are not unusual during this period. The program has been carefully tested and all the results are manually checked. Any ambiguous matches are verified using information from Dun & Bradstreet, company websites, as well as industry publications.

Patents assignees are first matched with the companies in the DCA file. Then the patent information is aggregated at the ultimate parent level according to the affiliation information from DCA. Accordingly, annual financial data for the American parents are obtained from the Compustat files and then matched with the consolidated patent files. When mapping multiple years together, I carefully checked the possible changes in company names, database coding, mergers and acquisitions, etc., to ensure consistency.

I drop those companies that have zero patent output during the entire sample period and those 3-digit SIC industries that contain less than three innovating firms. After data cleaning, the main sample contains 1567 firms in 92 three-digit SIC industries, whose patent output during the eight-year period ranges from one to over twenty thousand (IBM), averaging at more than one hundred per firm. Among these firms, 681 firms register positive utilization of foreign inventors while only 227 of them use inventors from weak IPR countries.

From the data sources I can 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 those innovations.

⁶ I thank Wilbur Chung for his help with the program.

4.5 Locations of Innovation

Empirical studies often take the country of the first author as the location of an invention, partly because this information is readily available in the NBER patent data. However, given my focus on human capital utilization, I want to be very careful about the possible distortion. In a patent with multiple inventors, which is often the case in cross-border R&D projects, an inventor from a developing country is likely to be the second, third, or 15th coauthor in the sequence. The first authors' 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 list⁷. 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 thought 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,

- 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 US, then this patent is considered to be developed in the home country; and
- The rest are technologies developed in other (strong IPR) foreign countries.

The 236,850 patents in the full sample are described in Table 1. Small as it is, the share of patents developed by American firms in weak IPR countries has been steadily increasing over the last decade.

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. I will vary the classification

⁷ There are no absolute rules as to the weighting. Basically, the weight is higher for inventors listed at the beginning of the sequence, and higher for inventors in a smaller team.

criteria to check the robustness of my results. Table 2 gives a glance to the 1567 firms included in the sample.

4.6 Internal Linkages

There is no direct measure for internalized value, but value can be proxied by usage. Technologies whose values are highly dependent on other internal resources are more likely to be utilized within the firm boundaries. Since “citations to patents that belong to the same firm” represent mostly internalized knowledge transfers (Hall, Jaffe and Trajtenberg 2001), I will use *self-citations* to proxy for the internalized value of each technology. Presumably, the more a patent is cited by the same firm (forward self citations), the more its value is being retained inside the firm boundary.

Note that my measure of “self-citation” differs from that of Hall, Jaffe and Trajtenberg (2001) in two important aspects, which lead to 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. For example, “Dell USA Corporation” and “Dell USA, L.P.” were treated as two different firms. 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 with the parent company or with one another. Because I am more interested in the firm as an integrated organization, any citation that occurs within the same firm boundary is considered a self-citation. For example, a citation from IBM Japan, Ltd. to a patent owned by the IBM in New York State would be counted as self-citation in this study, but would not be the case in HJT.

Similar to HJT, my citation calculation is subject to the truncation problem in the time series. The patents filed within 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 are likely to be affected 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, my measure will only favor those technologies that are cited by the same-firm patents faster. This is consistent with my objective to capture the efficiency in internal knowledge utilization: *speed* as well as *scale*.

V. MODEL AND RESULTS

So far I have obtained a rich set of variables to capture the theoretical concepts in the model. In this section I seek empirical support for the internalization-arbitrage conjecture by analyzing the within- and across-firm variations in technology structures.

5.1 Econometric Model

Let N_k be the number of citations received by technology k , among which n_k are self-citations. n_k , which proxies for internal linkages and falls within the interval $[0, N_k]$, depends on (1) the external IPR environment, (2) the internal firm characteristics.

I choose the zero-inflated negative binomial model to reflect three features of the data: First, most patents only received a small number of citations during the short sample period. A continuous measure such as self-citation ratio would be hard to justify where a large proportion of the observations are 0's and 1's. Second, negative binomial is preferred to a simple Poisson model due to the large variance in the number of received citations. Third, zero self-citations may arise from a different mechanism. Because many patents have yet to receive any citations till the end of the sample period, $n_k = 0$ does not necessarily mean a low level of internalization. It may simply be constrained by the fact that $N_k = 0$.

In the first step, I control for firm effect and study the within-firm variance. The regression takes the following form:

$$Pr(n_k) = f(\textit{weak}, \textit{foreign}, \textit{i.firm})$$

with *inflate* (N_k) *exposure* (N_k) *cluster* (*nclass*)

where

- *weak* is the dummy variable to indicate whether the technology is developed in a weak IPR country.
- *foreign* is the dummy variable to indicate whether the technology is developed in a foreign country. Hence, *foreign* must be 1 if *weak* = 1.
- *i.firm* is a series of dummies variables to control for firm fixed effects.
- *inflation* (N_k) specifies that the total number of citations N_k may cause the observed count n_k to be zero. *exposure* (N_k) specifies N_k as the maximum scope in which self-citations n_k can be observed for each observation.
- *cluster* (*nclass*) specifies that the observations are independent across patent classes but not necessarily within the classes. A generalized correlation matrix may be needed for the regression.

The coefficient on variable *foreign* is expected to be negative. Previous studies have found that knowledge diffusion is geographically concentrated in nature (Almeida 1996, Jaffe and Trajtenberg 1998). Therefore, I expect that the foreign developed patents are 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 appropriate the inexpensive human capital, and knowledge internalization is used as a barrier against imitation. Hence, the demand for 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, the firms may be developing similar technologies at multiple locations, but with different organizational structures. Both cases are consistent with the theoretical conjecture. Although a control for technology fields is not necessary from the theoretical perspective, 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 firms and examine whether firms doing R&D in weak IPR countries manifest stronger internal linkages. The regression takes the following form:

$$Pr(n_k) = f(\textit{weak}, \textit{foreign}, \textit{size}, f_weak, f_foreign)$$

with *inflate* (N_k) *exposure* (N_k) *cluster* ($nclass$)

where

- *weak* and *foreign* are defined as above.
- *size* is the firm's average assets or sales during the sample period. Alternatively, I also use total number of patents, and the logarithm of these variables, to proxy for firm size. Large firms, presumably, are more likely to see both ends of a citation falling within the firm boundary.
- *f_weak* and *f_foreign* are two dummy variables to indicate whether the firm that developed patent k has any R&D in weak IPR countries or any foreign countries.

The reason why the technology-specific variables *weak* and *foreign* remain in the regression is that R&D locations significantly affect the degree of internalization. It would be hard to compare across firms without controlling for the location differences. For example, a patent developed by an American firm in Germany would be hardly comparable to a patent developed by an American firm in the US. By controlling for the location effect, I am able to examine whether, among technologies developed under similar environment, a firm's R&D presence in weak IPR countries is associated with stronger internalization of its technologies.

According to the discussion in previous sections, firms need to have strong organizational capabilities to successfully implement R&D internalization, even more so for R&D in weak IPR countries. Therefore, I expect positive coefficients on both variables *f_weak* and *f_foreign*.

Again, the difference in internal linkages may arise from firms' heterogeneous organizational capabilities. It may also arise from the particular technology fields that the firms are involved in. I cannot precisely tell these two cases apart, 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 knowledge structure and its multinational R&D strategies.

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

$$\frac{\partial E(n_k)/N_k}{\partial x_i} = \frac{E(n_k)}{N_k} \cdot \beta_i$$

Because the key independent variables are 0-1 dummies, the coefficients can be roughly interpreted as the change in self-citation ratio if the dummy changes from 0 to 1.

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 developed under different IPR regimes. Table 3 gives the 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 regression results are shown in Table 4. Column (1) is the base model run on the whole sample. With the truncated data, self-citation ratio is increasing over time because inventors tend to cite the same-firm patents faster. In column (2) I repeat the same regression with eight year-dummies to remove the time effect. Replacing the year-dummies with a trend variable t generates similar results.

In column (3), 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, those above certain R&D scales should be the more appropriate group to study.

Statistics show that over half of the US patent citations are imposed by examiners, rather than being submitted by inventors or patent attorneys in their patent applications (Alcacer

and Gittelman, 2003).⁸ 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. Therefore in column (4), I conduct the test using only inventor-filed citations made by post-2000 patents.

In the theoretical analysis, I argue that the viability of this strategy may vary widely across industries. When the R&D needs substantial support from the local knowledge base, the low wages may be compromised by low efficiencies. It would be interesting to examine whether the internalization-arbitrage conjecture holds not only in aggregate, but also in particular technology fields. Column (5) reports the results for patents in the “Computers and Communications” field, as defined by Hall, Jaffe and Trajtenberg (2001). It contains 35 primary US patent classes spanning communications, computer hardware & software, computer peripherals, and information storages. Separate tests are also conducted in other categories.

As shown in Table 4, throughout the specifications the coefficient on the *foreign* dummy is significantly negative, which indicates distance effect. Marginal effect calculation shows that 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 IT industry and pharmaceutical industry.

However, the positive coefficient on the *weak* dummy nearly offset 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 US headquarters!

It may be the case that the large proportion of self-citations is not due to firms’ strategic project assignment, but to the development level of the host countries. Generally, the weak IPR countries are also the less developed countries. Innovators in those countries

⁸ I thank Juan Alcacer and Michelle Gittelman for their help on data preparation.

have to rely heavily on the knowledge base of the parent companies and produce very firm-specific intellectual properties. However, this interpretation is still consistent with the theoretical argument. Multinationals provide the knowledge base necessarily to develop the new technologies effectively, which cannot be easily obtained otherwise by the local talents. As a result, the generated technologies are less likely to be utilized outside of the firm boundaries.

5.3 Across Firms: Organizational Structure

In this step I look across firms and examine firm capabilities. Do firms that conduct R&D in weak IPR countries systematically differ from other firms in their internal 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 on average.

The regression results are shown in Table 6. Column (1) and (2) are the base model run on the whole sample, with and without time effect. In column (3), I control for patent classes⁹ to test whether the difference in internal linkages are driven by different technology fields that the firms are engaged in. Same as in the within-firm analysis, column (4) removes examiner-imposed citations and focuses on citations received after 2000, while column (5) gives a glimpse of the results for specific technology fields.

Across all specifications, the coefficients on f_weak and $f_foreign$ are positive and significant. The results show that firms who are able to 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 are able to do R&D in weak IPR countries manifest even stronger internal linkages. The location effects are controlled throughout the analysis.

⁹ The reported result uses the 3-digit US primary patent classes. Aware of the caveats with this classification, I also used the International Patent Classes (IPC) on a sub-sample and obtained qualitatively similar results.

Arguably, firms with R&D in weak IPR countries have proportionally more self-citations simply because they are larger. Although this is consistent with my theory¹⁰, 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 with controls of assets (1), the logarithm of assets (2), annual sales (3), and the total number of patents owned by the firm (4).

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 very small. 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 level of self-citations, but the economic significance is, on average, less than 1% of the f_weak and $f_foreign$ dummies. 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.

Bias in Using Patent Data

Without a direct measure for knowledge creation and knowledge flows, there are potential measurement biases in the empirical test:

- I hope to test whether *technologies* developed in weak IPR countries are more internalized than technologies developed elsewhere. But instead we can only observe whether *patents* developed in weak IPR countries receive higher self-citations.
- I hope to test whether firms that *do R&D* in weak IPR countries have more internalized innovation structures. But instead we can only observe whether firms that *have patents* from weak IPR countries are more internalized.

¹⁰ Optimally, firms expand until the organizational costs go beyond the benefit of internalization. In that sense, firm size itself reflects organizational capabilities.

With different patent propensity across firms and across countries, I need to examine to which extent the unobservable part of the R&D activities would affect my result.

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. Moreover, the results of many R&D activities, such as program coding and laboratory tests, are not patentable at all. These activities also tend to be the most firm specific. Therefore, the non-patented technologies should be more internalized than the patented ones. 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 China labs suggest that most R&D efforts there are not in the patent-oriented category. Only a small proportion of the technologies with the highest outside value will be patented in the US. 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 by 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.

Biases in Measuring Self-Citations

Due to the organizational changes over the years, the self-citation counts may be biased upward or downward. For example, after firm *A* was 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 assigned to *A* would not be counted as self-citations, even if the citing and cited patents may involve exactly the same team of inventors. On the other hand, if a team in firm *A* worked on a project up to 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 within-firm knowledge flows.

To avoid these biases, I eliminate from the sample any assignees that changed firm affiliation during the sample period. Also eliminated are firms that substantially changed their names, although the affiliation numbers remain the same. This decreased the number of firms in my sample to 1054.

The same within-firm and across-firm regressions are conducted on the reduced sample, and the results from the base model are reported in Table 8. Compared with the base model in Table 4 and Table 6, the reduced sample with stable firm affiliations produces even stronger within- and across-firm differences. Controlling for firm effect, technologies developed in weak IPR countries are 25% more likely to be cited internally than those developed in other foreign countries. Controlling for R&D locations, firms with R&D in weak IPR countries have roughly 43% higher self-citation ratios than firms that only engage in R&D in strong IPR countries, and 74% higher than purely domestic firms.

Alternative Measure of Internal Linkages

There are many aspects of internal linkages for an organization, and the self-citation measure is at best an imperfect proxy. The internalization-arbitrage argument would be more convincing the same pattern can be found 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.

Again, I follow the same procedures of within- and across-firm analysis.

Suppose in each firm, there are K_c patents involving inventors from a particular country c , among which k_c are collaborated work with inventors from other countries. Then the ratio k_c/K_c captures the linkages between the inventors in the country c subsidiary and other parts of the firm. Focusing on firms with R&D in weak IPR countries, I find that on average, 62% of the patents involving inventors from weak IPR countries are

collaborated, versus 44% for patents involving inventors from other foreign countries. Among patents that are developed at 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 firms staying away.

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 overall.

VII. CONCLUSION AND FUTURE EXTENSIONS

Doing R&D in countries with weak IPR protection seems contradictory to the comparative advantage theory. Because of the poor institutional environment, these countries are not known for their R&D or technology strengths. This paper argues that multinational firms can substitute their internal innovation organizations for the external environment. Firms with closely-knit internal technology structures are able to take advantage of the undervalued human capital in these 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 infrastructure matters.

While identifying the potential opportunities in adverse institutional environment, 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 strategically internalize their knowledge-intensive activities. As a result, even if the quantity of foreign R&D may not be significantly affected by patent protection (Kumar 2001), the nature of R&D varies widely, as evidenced by this study. In particular, a multinational firm would only carry out specific types of R&D in weak IPR countries, and only firms with strong organizational capabilities would choose to do so. In other words, internalization-arbitrage is multinational firms' strategic choice under the constraint of poor external environment. Removing this constraint is expected to alter both the channel and the direction of knowledge flows. For example, multinationals would feel comfortable bringing technologies that are more readily applicable to the host countries. Indeed, Branstetter, Fisman and Foley (2002) show that U.S. multinational firms responded to IPR reforms in the host country with more technology transfers to the subsidiaries.

For policy makers, this study points out that firms are not passive policy takers. Multinational 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. The interactions between firm strategies and institutional changes have to be analyzed in a global context.

This study also opens a whole line of further researches.

First, it would be interesting to take the current analysis to a dynamic setting. R&D internalization and R&D in weak IPR countries have both grown rapidly in the recent several years. What are the fundamental changes that have magnified this arbitrage-internalization effect? Are these changes happening in corporate management or in the external environment? Or both? Or one is the repercussion of the other?

Second, even among firms that do R&D in weak IPR countries, I have observed a large variation in R&D organizations across firms and across industries. How is this related to firm performance? Is this dependent on the leader/follower positions of individual firms? Answering these questions entails in-depth study to disentangle the multiple forces underlying the variation.

Finally, the arbitrage-internalization mechanism should not be limited to innovation only. This thinking – using internal capability to substitute for inadequate social and legal institution – can also be found in the mainstream internalization theories (e.g., in Buckley and Casson, 1976) and in the line of work on “business groups” (Khanna 2000). Although the intangibility of knowledge makes R&D the ideal sector for institutional arbitrage, we have reason to believe that the mechanism applies to other sectors too. A study with a boarder view would surely deepen our understanding of firms and institutions.

BIBLIOGRAPHY

- Alcacer, Juan & 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.
- Ansberry, Clare. 2003. Outsourcing Abroad Draws Debate at Home. Wall Street Journal, The Outlook, July 14, 2003
- Bartlett C. A., & Ghoshal S. 1990. The Multinational Corporation as an Interorganizational Network. Academy of Management Review, 15(4): 603-625
- Branstetter, Fisman and 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. & Mark Casson, C. 1976, 2002. The Future of the Multinational Enterprise. (1st edition, 1976, London: Macmillan, 2nd edition, 2002, Palgrave).
- Cantwell, John & 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 & Juan Alcacer. 2002. Knowledge Seeking and Location Choice of Foreign Direct Investment in the United States. Management Science, 48(12): 1534-54.
- Cockburn, Iain & 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. & Daniel Levinthal. 1990. Absorptive Capacity: A New Perspective on Learning and Innovation. Administrative Science Quarterly. 35: 128-152.
- Cohen, Wesley M. & Steven Klepper. 1996. A Reprise of Size and R&D. The Economic Journal, 106(437): 925-51.
- Cohen, Wesley M., Richard R. Nelson & 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 & 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. Forthcoming.
- 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, & Manuel Trajtenberg. 2001. The NBER Patent Citation Data File: Lessons, Insights and Methodological Tools. NBER Working Paper, No. 8498.
- 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, W. 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 & 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 & 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 & Pablo Zoido-Lobaton. 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, & Krishna G. Palepu. 1997. Why Focused Strategies May Be Wrong for Emerging Markets. Harvard Business Review, 75(4): 41-51.

- Kogut, Bruce & 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 US 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. Work in progress.
- Lee, Jeong-Yeon & Edwin Mansfield. 1996. Intellectual Property Protection and U.S. Foreign Direct Investment. The Review of Economics and Statistics, 78(2): 181-86.
- Mansfield, E, M. Schwartz & S. Wagner. 1981. Imitation Costs and Patents: an Empirical Study. Economic Journal, 91: 907-18.
- Mitchell, Will, Randall Morck, Myles Shaver & Bernard Yeung. 1998. Causality between International Expansion and Investment in Intangibles, with Implications for Financial Performance and Firm Survival,” in Global Competition and Market Entry Strategies, J-F Hennert (ed.), Elsevier, North-Holland, 1998
- Morck, Randall & Bernard Yeung. 1991. “Why investors value multinationality,” Journal of Business, 64(2): 165-87.
- Patel, Pari & 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 & Modesto Vega. 1999. Patterns of internationalisation of corporate technology: Location vs. home country advantages. Research Policy, 28(2-3): 145-55.
- Rapp, Richard T. & Richard P. Rozek. 1990. Benefits and Costs of Intellectual Property Protection in Developing Countries. Journal of World Trade, 75: 78-81.
- Teece, David J. 1986. Profiting from Technological Innovation: Implications for Integration, Collaboration, Licensing and Public Policy. Research Policy, 15(6): 285-305

TABLES

Table 1. Description of 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	251	0.81%	6,651	21.41%	31,061
1996	278	0.95%	4,387	14.94%	29,371
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	2,836	1.20%	30,070	12.70%	236,850

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)	10414.65	4981.98	2687.09
Sales (million dollars)	8088.27	3813.81	1931.97
Number of patents	750.14	298.60	137.61
Number of subsidiaries	45.69	31.99	19.36
Number of assignees	4.98	3.34	2.27
Countries with presence	13.56	8.92	4.55
H-index for tech class	0.08	0.12	0.28

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

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	0.133	0.191	0.129	0.228	0.083	0.198
1994	0.167	0.252	0.130	0.232	0.076	0.192
1995	0.187	0.267	0.117	0.231	0.070	0.180
1996	0.198	0.302	0.109	0.237	0.078	0.215
1997	0.181	0.295	0.140	0.277	0.101	0.247
1998	0.189	0.314	0.179	0.330	0.166	0.324
1999	0.236	0.377	0.216	0.360	0.212	0.357
2000	0.399	0.472	0.251	0.398	0.365	0.453
2001	0.686	0.464	0.315	0.449	0.396	0.455
Average	0.200	0.312	0.164	0.301	0.224	0.324

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

	Without Year Dummy (1)	With Year Dummy (2)	Firms with >50 patents (3)	Excluding examiners' (4)	Computer & Telecom (5)
Weak IPR country	0.2201^{***} (0.0429)	0.2055^{***} (0.0425)	0.2033^{***} (0.0426)	0.2128^{***} (0.0566)	0.3941^{***} (0.0659)
Foreign country	-0.2851 ^{***} (0.0152)	-0.2838 ^{***} (0.0150)	-0.2838 ^{***} (0.0151)	-0.2248 ^{***} (0.0217)	-0.4496 ^{***} (0.0303)
Const	-1.6529 ^{***} (0.0595)	-1.6743 ^{***} (0.0599)	-1.6755 ^{***} (0.0600)	-1.9263 ^{***} (0.1093)	-0.6564 ^{***} (0.6398)
Total citation	- exposure				
Inflate					
Total citation	0.0131 ^{***} (0.0011)	0.01181 ^{***} (0.0012)	0.01180 ^{***} (0.0022)	0.0883 ^{***} (0.0078)	0.0178 ^{***} (0.0022)
Const	-2.4702 ^{***} (0.0484)	-2.4389 ^{***} (0.0497)	-2.4405 ^{***} (0.0498)	-2.9442 ^{***} (0.1213)	-3.2100 ^{***} (0.1309)
Obs	125796	125796	125036	95302	42801
log_likelihood	-153830.20	-153479.8	-153106.1	-66659.79	-52536.13
LR chi2	22600.59	23301.52	22799.91	14710.83	7282.7
Prob > chi2	0.00	0.00	0.00	0.00	0.00

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

	Without Year Dummy (1)	With Year Dummy (2)	Control for patent classes (3)	Excluding examiners' (4)	Computer & Telecom (5)
Firm w/ R&D in weak IPR	0.3991^{***} (0.0127)	0.3950^{***} (0.0125)	0.3239^{***} (0.0126)	0.5337^{***} (0.0191)	0.5256^{***} (0.0249)
Firms w/ R&D in foreign	0.2644 ^{***} (0.0290)	0.3196 ^{***} (0.0285)	0.3112 ^{***} (0.0284)	0.1016 ^{**} (0.0403)	0.7188 ^{***} (0.0652)
Developed in weak IPR	0.0808 [*] (0.0467)	0.0390 (0.0456)	0.2002 ^{***} (0.0438)	0.0573 (0.0585)	0.4846 ^{***} (0.0683)
Developed in foreign	-0.4077 ^{***} (0.0146)	-0.3811 ^{***} (0.0144)	-0.4605 ^{***} (0.0143)	-0.2753 ^{***} (0.0206)	-0.6570 ^{***} (0.0282)
Const	-2.0770 ^{***} (0.0272)	-2.2373 ^{***} (0.0286)	-2.5436 ^{***} (0.4744)	-2.8044 ^{***} (0.0446)	-3.0136 ^{***} (0.0650)
Total citation	- exposure				
inflate					
Total citation	0.0119 ^{***} (0.0009)	0.0102 ^{***} (0.0010)	0.0097 ^{***} (0.0009)	0.0830 ^{***} (0.0052)	0.0145 ^{***} (0.0019)
Const	-2.0326 ^{***} (0.0455)	-2.0344 ^{***} (0.0492)	-1.9433 ^{***} (0.0409)	-1.9950 ^{***} (0.0778)	-2.7041 ^{***} (0.1306)
Obs	153950	153950	153950	116138	49733
log_likelihood	-192640.3	-191527.2	-186562.4	-83201.8	-62134.3
LR chi2	2180.04	4406.28	14335.92	5538.83	2044.86
Prob > chi2	0.00	0.00	0.00	0.00	0.00

Table 7. Cross-Firm Regression with Size Effect

	Control for Assets (1)	Control for Log of Assets (2)	Control for Sales (3)	Control for Patent Output (4)
Assets (\$BN)	-0.0020*** (0.0001)			
Log_Assets		-0.0219*** (0.0034)		
Sales (\$BN)			-0.0024*** (0.0002)	
Patent Output (thousand)				0.0120*** (0.0007)
Firms w/ R&D in weak IPR	0.4761*** (0.0141)	0.2323*** (0.0165)	0.4788*** (0.0142)	0.1095*** (0.0146)
Firms w/ R&D in foreign	0.3415*** (0.0326)	0.2878*** (0.0338)	0.3422*** (0.0326)	0.2756*** (0.0302)
Developed in weak IPR	0.0340 (0.0471)	0.0672 (0.0476)	0.0410 (0.0471)	0.0980* (0.0473)
Developed in foreign	-0.3655*** (0.0154)	-0.3718*** (0.0154)	-0.3686*** (0.0154)	-0.4043*** (0.0151)
Const	-2.2574*** (0.0326)	-2.9459*** (0.0902)	-2.2638*** (0.0326)	-3.1274*** (0.0827)
Total citation	- exposure			
inflate				
Total citation	0.0089*** (0.0012)	0.0107*** (0.0012)	0.0088*** (0.0012)	0.0135*** (0.0012)
Const	-2.1412*** (0.0326)	-2.2904*** (0.0589)	-2.1321*** (0.0578)	-2.5187*** (0.0696)
Obs	126636	125672	126636	137626
log_likelihood	-160355.5	-159688.5	-160369.4	-167770.0
LR chi2	3766.21	3357.86	3738.43	7726.48
Prob > chi2	0.00	0.00	0.00	0.00

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

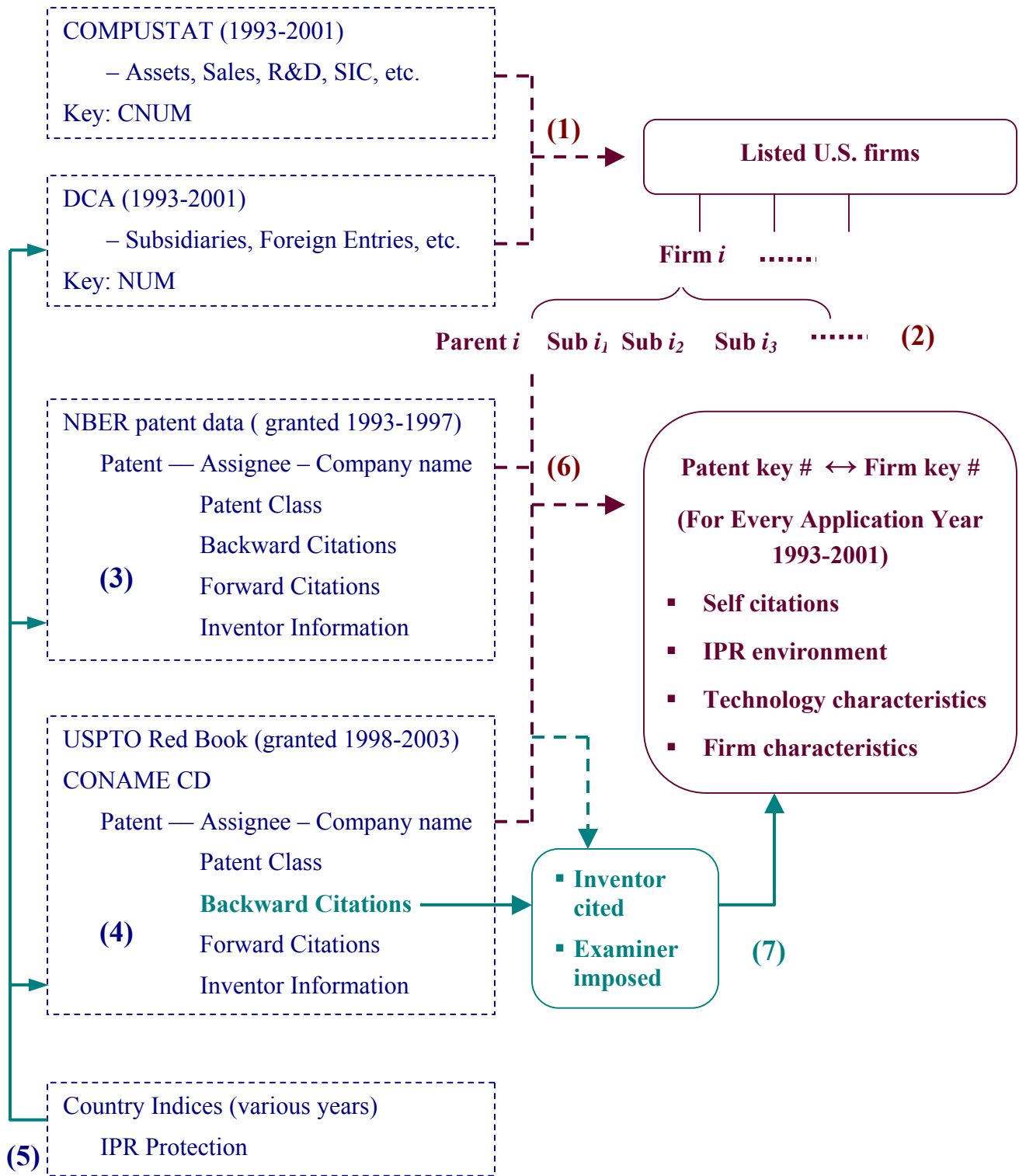
	Within-firm (1)	Across-firm (2)
Firms w/ R&D in weak IPR		0.4309^{***} (0.0136)
Firms w/ R&D in foreign		0.3115 ^{***} (0.0314)
Developed in weak IPR	0.2543^{***} (0.0437)	0.0826 [*] (0.0471)
Developed in foreign	-0.3034 ^{***} (0.0159)	-0.4131 ^{***} (0.0152)
Const	-1.6456 ^{***} (0.1396)	-2.2466 ^{***} (0.0315)
Total citation - exposure		
inflate		
Total citation	0.0105 ^{***} (0.0014)	0.0095 ^{***} (0.0012)
Const	-2.5555 ^{***} (0.0589)	-2.1863 ^{***} (0.0610)
Obs	106365	129614
log_likelihood	-130730	-162695.3
LR chi2	20209.91	3855.15
Prob > chi2	0.00	0.00

APPENDIX A. Countries of Weak IPR Protections

	Scientist &Enginee rs/million people	Population (millions)	Tertiary School (gross %)	Opacity Factor 2000	Property Rights 1995	Law & Order 93-95	Rapp and Rozek	Ginarte & Park 1990	USTR's Special 301 ¹¹ 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	1215.30	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
HK, 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 (CN)	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

¹¹ 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 mid-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.