Patent Pools, Thickets, and the Open Source Commercialization Strategies of Start-Up Firms [*]

Wen Wen
College of Management
Georgia Institute of Technology
800 West Peachtree St. NW
Atlanta, GA 30308
wen.wen@mgt.gatech.edu

Marco Ceccagnoli
College of Management
Georgia Institute of Technology
800 West Peachtree St. NW
Atlanta, GA 30308
marco.ceccagnoli@mgt.gatech.edu

Chris Forman
College of Management
Georgia Institute of Technology
800 West Peachtree St. NW
Atlanta, GA 30308
chris.forman@mgt.gatech.edu

April 2011

PRELIMINARY AND INCOMPLETE – PLEASE DO NOT QUOTE OR CITE WITHOUT PERMISSION

Abstract

How did the introduction of open source patent pools affect the open source commercialization strategies of start-up firms? We examine the relationship between the size of open source patent pools and the entry of start-up software firms employing an open source licensing strategy between 1999 and 2009. We find that increases in the number of patents in one pool—the Patent Commons—is associated with increased open source entry. However, increases in the number of patents in a separate pool—the Open Innovation Network—is not associated with any increase in entry. Patents in both pools have a higher number of average forward citations, but a lower number of backward citations and lower number of claims.

Keywords: open source, patent pools, intellectual property strategy
JEL Classification: O34, L86

[*] We gratefully acknowledge funding from a Kauffman Foundation/Georgia Research Alliance grant for the study of entrepreneurship and productivity. Wen Wen thanks the Kauffman Foundation for providing funding for this research through a Kauffman Dissertation Fellowship. Chris Forman acknowledges funding from the Alfred P. Sloan Foundation through an Industry Studies Fellowship. All errors are our own.

1. INTRODUCTION

An increasing body of theoretical and empirical work has highlighted the potential costs of the patent system to cumulative innovation. In many technological fields there exist significant patent thickets, a dense web of intellectual property rights that firms must navigate their way through in order to commercialize new technology. These issues are likely to be particularly salient for technologies in which standards and standardization play an important role: in such fields, firms producing to the standard will often need to license complementary technologies from a range of firms. As has been noted elsewhere, in such settings this range of potential blocking patents can give rise to a tragedy of the anti-commons (Heller and Eisenberg 1998). A range of mechanisms are available for addressing this issue, including cross licenses and a cooperative standard-setting process that requires participants to license essential patents on reasonable terms (Shapiro 2000). Another potential mechanism is patent pools—institutions under which volunteer organizations license the patents of two or more companies to third parties in a single package. Patent pools are emerging as an increasingly important area for economic analysis, but there has been little empirical work studying the implications of patent pools for the direction of innovative activity.[1]

In this paper we seek to study how the introduction of open source patent pools influences a key commercialization strategy of start-up software firms: the decision to license their software as open source. In particular, we examine how the density of patent thickets and the introduction of open source software (OSS) patent pools interact to determine the rate of entry of new OSS products from start-up firms within different software markets. We study this particular commercialization strategy for several reasons. First, it is a relevant decision for many

[1] For one important exception, see Lampe and Moser (2009). For other examples of recent empirical work on patent pools, see Layne-Farrar and Lerner (2010), Gilbert (2004), and Lerner, Strojwas, and Tirole (2007).

start-up firms: as noted by Lerner and Schankerman (2010), open source and proprietary code are frequently comingled and many firms develop both types of software. Second, a variety of anecdotal evidence has suggested that the costs of patent thickets and licensing complementary innovation may be particularly high for OSS firms. The free flow of source code among OSS developers implies a higher dependence on cumulative innovation than traditional proprietary software, leading to potentially higher costs of licensing complementary technologies (Shapiro 2000). Moreover, the existence of many anonymous individual developers makes it difficult to identify the provenance of source code, imposing high costs for OSS to identify potentially infringing technologies. Also, as noted by Graham and Mowery (2005), OSS may be particularly susceptible to "submarine patent" strategies.

While patent pools are thought to ameliorate the potential anti-commons problem, to our knowledge there is no prior empirical research that tests the impact of OSS patent pools on the OSS commercialization strategies of start-up firms. Unlike a traditional patent pool that licenses a set of patents to each member of the pool or to third parties and specifies the rule of royalty earnings among members, OSS patent pools in our setting offer royalty-free usage of all patents to any OSS firms that promise not to sue the pools' beneficiaries. However, the major benefits of OSS patent pools to start-ups and contributors are likely to be similar to that of traditional patent pools. For contributors, patent pools may promote complementary innovation that increases the value of contributor's assets. For start-up firms, pools provide a set of defensive patents that may ameliorate the anti-commons problem. Empirical evidence from this study may not only demonstrate the impact of patent pools on the rate of innovation in the software industry, but can also help address questions about the value of such pools.

Motivated by these observations, our specific research question is to investigate the impact of OSS patent pools on the rate of new OSS product introduction by software startups. In particular, we hypothesize that (i) that the number of contributed patents in the pool related to a software market segment will induce entry with new products into the OSS regime by start-ups in that segment and (ii) that the impact of patent pools on start-up entry into the OSS industry will be greater in market segments with denser patent thickets.

We examine the empirical salience of these hypotheses using a unique data set combining information from a variety of sources. We assemble data on new open source product introductions from the news releases of start-up software firms announced in the PROMT database. Following prior work that has examined the extent to which patents deter entry in the software industry (Cockburn and MacGarvie 2009, 2011), we map patent technological classes to their most similar software product market segment. We use this to identify the stock of patents by technological class in each market segment as well as the number of pool patents. Our empirical strategy examines whether growth in the number of pool patents within market segments over time is associated with increases in market entry by open source software firms.

At present our results are at a preliminary stage. We show that, on average, increases in the number of pool patents are associated with an increase in open source software (OSS) market entry. These results are controlling for product market segment and time fixed effects. However, we show that these results are disproportionately due to one of the two OSS patent pools. Further work will examine how the marginal effect of patent pools varies based upon the density of the patent thicket in the product market. Further, we will examine in greater detail the extent to which omitted variables in our regressions may influence our results. We discuss in further detail our plans for addressing these issues below.

2. INSTITUTIONAL BACKGROUND

Two major OSS patent pools were established recently with the central goal to provide the OSS community a body of patents to freely draw on for cross-licensing arrangements. The first pool is the Open Source Development Lab's Patent Commons project (denoted as "the Commons" thereafter), which was established by IBM in early 2005. In January 2005, IBM pledged open access to its more than 500 software patents to "any individual, community, or company working on or using software that meets the Open Source Initiative (OSI) definition of open source software now or in the future." Similar to the pledge practice IBM have initiated in other fields, e.g. the creation of Eco-Patent Commons to protect green technologies (Hall and Helmers 2011), "pledge" in this context means the pledged patents will be offered royalty-free to any third parties that are engaging in activities that might otherwise give rise to a claim of patent infringement and that promise not to sue the pools' beneficiaries. [2] Thus, OSS participants can freely embed technology covered by these patents into their own software without any fear of being sued. As announced in the press release, IBM believed this was the largest pledge ever of patents of any kind. It is worth noting that one major difference between the Commons and other traditional patent pools is that no formal agreement is needed between the Commons and the beneficiary of the pool to use these patents. Meanwhile, to safeguard the appropriate use of pledged patents, the Commons explicitly suggested an additional goal to "discourage lawsuits from being brought against beneficiaries of the pledges by making clear the covenant not to sue does not apply to those who sue a beneficiary for patent infringement or, in some commitments, on any intellectual property claim."[3]

---

[2] For more details, see http://www.patent-commons.org/resources/about_commitments.php.

[3] http://www.patent-commons.org/about/the_commons.php

The other OSS patent pool of our focus is the Open Innovation Network (denoted as "OIN" hereafter). Since late 2005, OIN has been accumulating a set of patents from purchase, auction, and donation. It offers contractually royalty-free usage of these patents to OSS participants for use in their own products as long as the patent users make no future patent infringement claim against the Linux System[4]associated software. Consistent with the goal of the Commons, OIN has a stated goal to establish a collaborative environment to free up the flow of technology related to Linux and to reduce the OSS community's fear of lawsuits from patent claims. Nevertheless, there are a few aspects that make OIN slightly different from the Commons. First, while the Commons does not require a formal contractual agreement for the beneficiary to use patents in the pool, the OIN requires a formal license agreement[5] in which any licensee agrees not to assert their own patents against Linux and Linux-related systems and applications. Second, while most of the patents in the Commons were pledged by IBM, OIN is a formal company founded in November 2005 by IBM, Novell, Philips, Red Hat, and Sony. Therefore, if the motivation for large incumbents to pledge patents or fund a patent pool is to spur open innovation complementary to their core business, we would expect these two pools may have distinct impacts on a variety of OSS segments, as OIN has contributors who operate in different technology areas. In addition, the time pattern of contributions differs across the two pools. IBM contributed its more than 500 patents to the Commons all at once (in early 2005); since then, there have been few contributions to the Commons. On the other hand, the cumulative stock of patents available through OIN has been changing over time, as shown in

[4] Linux system mean "a Linux Environment Component or any combination of such components to the extent each such component is (i) generally available under an Open Source License or in the public domain (and the source code for such component is generally available) and (ii) Distributed with, or for use with, the Linux Kernel (or is the Linux Kernel)." See http://www.openinventionnetwork.com/pat_linuxdef.php .

[5] A detailed license agreement is available on http://www.openinventionnetwork.com/pat_license_agreement.php.

Figure 1. As a result of these differences, our empirical analysis will focus on both the overall and the separate impact of the two OSS pools.

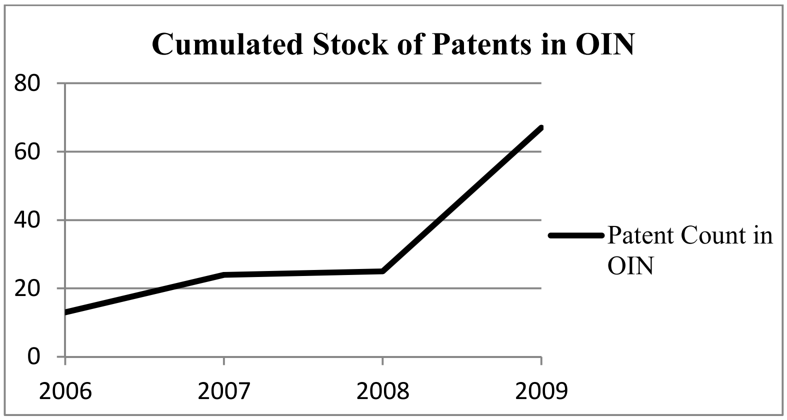

Figure 1: The number of patents in OIN (raw data)

The potential benefits from these OSS patent pools to small OSS firms are expected to be substantial, since firms do not have to specifically request a license to any of more than 600 patents in both OSS pools, which grants a more convenient but less costly access to a variety of technologies. As noted by Matt Asay, the chief operating officer at Canonical (the company behind the Ubuntu Linux operating system), "this (type of patent collectives) may be the only refuge for start-ups and others, like Red Hat, that don't have an aggressive patent-acquisition policy" (Asay 2010). Nevertheless, some controversial views also arose from observers, some of whom are skeptical about the effectiveness of such defensive patent pools. One comment, for instance, was "the perception is that bigger companies only commit their least-effective, least-important patents to a patent pool" (Seeker 2010). Meanwhile, the extent of the heterogeneity between the Commons and OIN is also not clear-cut. Sorting out these puzzling issues is important, as it will shed light on the underlying motivations for large incumbents to open up some technology space for OSS entrants and help with the identification of how the two pools influence small entrants' commercialization of OSS across different software segments.

3. THEORETICAL BACKGROUND

There are several mechanisms under which the establishment of OSS patent pools related to a product market segment may reduce entry costs and encourage start-ups to develop new OSS products in that segment. First, the OSS patent pools provide start-ups with royalty-free usage of patents, so this could significantly reduce the costs of licensing (or the cost to invent around) existing technologies. This is particularly important in industries where subsequent innovations have a high degree of "cumulativeness" (Scotchmer (2004)). In the absence of such pools, royalties required by existing patentees may deter entry. The establishment of OSS patent pools may make start-ups aware of the existence of some patents that could potentially be used as substitutes for the blocking patents. Further, as the number of the pools' patents related to a market segment increases, it increases the likelihood that these patents can be used as substitute patents by start-ups to accumulate the necessary technological capabilities for entry. In sum, we argue that the pools' patents could reduce start-ups' innovation costs by serving as substitutes for the existing blocking patents in one market product segment, which may facilitate start-up's entry into that segment.

Second, consistent with the literature highlighting the strategic uses of patents (e.g. Cohen et al. (2000)), OSS patent pools may strengthen a start-up's negotiation position when bargaining over other blocking patents that are not part of the pool and may be held by nonparticipants in the patent pool. This is because OSS patent pool licensees can leverage patents in the pool to defend themselves against other potentially blocking patents. For instance, patents in the pools could be used for review and submission of prior art if a start-up is sued by firms holding blocking patents. This in fact was the case when OIN helped TomTom to defend itself against the allegation by Microsoft that TomTom was infringing on some of its patents.[6] Similarly, the Commons also established a partnership with the USPTO to ensure that patent examiners have access to all available prior art in the pool relating to the patent in question.[7]

Finally, as highlighted in prior literature (e.g. Cockburn and MacGarvie (2009, 2011)), in a software segment with denser patent thickets, entrants need to negotiate over a larger set of blocking patents, the costs of which are determined by both the entrant's negotiation power and number of blocking patents in the segment (e.g. Shapiro (2000), Ziedonis (2004)). Thus, the impact of patent pools on strengthening start-up's bargaining power should be greater in segments with denser patent thickets, since the pools can help start-ups to further reduce costs in negotiating over blocking patents.

4. DATA

4.1. Identify Start-up Entry into OSS across Software Segments

Our sample consists of a longitudinal dataset from 1999 to 2009 of 1200 US software firms from the CorpTech Directory 2010 of Technology firms.[8] The 1200 firms are all founded after 1990 and have either fewer than 1000 employees or less than $500 million in annual sales. Our research design requires us to identify new OSS startup entry and related patents across a range of market segments. Following Fosfuri et al. (2008), our first step is to identify start-up entry into

[6] http://www.openinventionnetwork.com/press_release04_28_09.php

[7] http://www.patent-commons.org/news/index.php?displaynews=17&page=1

[8] Our choice of 2010 CorpTech data reflects a constraint with the data—we have contacted CorpTech and there are no historical data between 2005-2009, the core years of our sample period. We acknowledge the potential for survival bias from this choice of data, and are currently working to extend our sample based upon a sample of data from CorpTech 2004 that we have obtained.

OSS regimes by searching the press releases of the 1200 software firms in the PROMPT[9] database. After downloading all news articles related to the focal 1200 firms from 1999 to 2009, we implement the following two step algorithm to determine new OSS startup entry. First, following work by Fosfuri et al. (2008) and Bessen and Hunt (2007), we searched for a set of keywords within PROMT articles to help us identify articles related to open source software. Some examples of these keywords include open source, OSS, FLOSS, General Public License (GPL), GNU, Lesser General Public License (LGPL), FreeBSD, Apache License, copyleft. Appendix 1 includes the full set of keywords. After manually reading a sample of these articles, we believe our identification method of OSS is not too-inclusive such that most of the search results are directly related to some software licensed as open source. Second, from the search results, we further identified the product introduction events by hiring a group of undergraduate students to manually read each articles. These two steps results in about 550 OSS product introduction events by 150 start-ups from 1999 to 2009. The figure 2 below presents the raw number of OSS product introduction events by start-ups from 1999 to 2009.

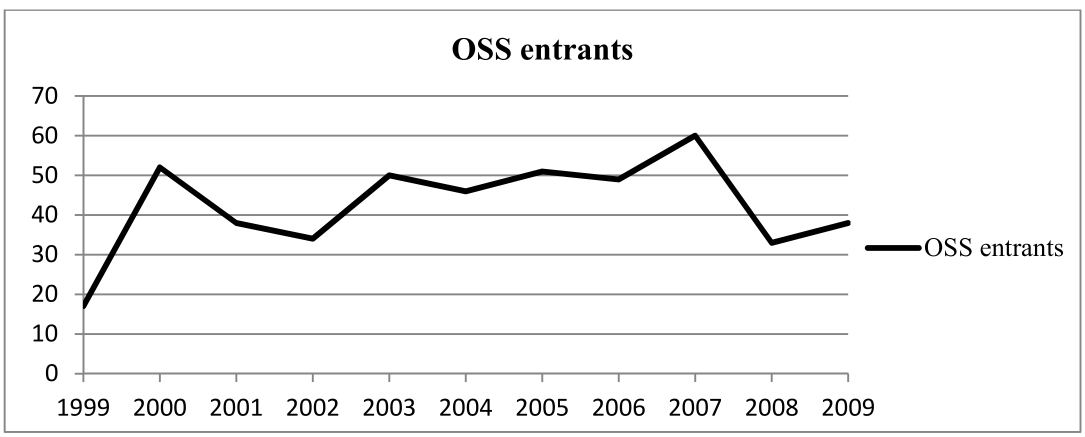

Figure 2: OSS product introduction events by start-ups from 1999 to 2009

[9] PROMT offers comprehensive coverage of companies, products, markets, and technologies from a vast collection of journals, newsletters, news releases and newspapers.

To measure the number of OSS product introduction events in different product market segments in each year, a crucial step is to divide the software market into different segments that are reasonably distinctive from each other. Our primary source of software segments is the product code classification system embedded in the PROMT database. These product codes are included in PROMT articles to indicate which product category/categories the article is associated with. All these product categories are organized in a hierarchical structure by PROMPT and are defined both in terms of customer segments and technologies. Appendix 2 provides some examples of PROMPT codes.

As we describe in further detail below, an important part of our data construction involves matching product market segments to patents—both to identify the density of the patent thicket that firms must navigate and also to identify product market segments where patents in the Commons and OIN pools are most likely to have an impact. Following work by Cockburn and MacGarvie (2009, 2011), we match software patents to CorpTech "SOF" product classes. To facilitate our mapping between software products and patents, we further consolidate PROMT product code classes with CorpTech product code classes[10]. The final classification that we use in our regressions consists of about 170 software segments (denoted as PROMPT-CorpTech classification hereafter). Appendix 3 shows some examples of the classes.

An additional data issue that we must confront is that some portion of news articles from PROMPT has its product code field missing. These articles account for 40% of the news articles identified as OSS product introduction events. We are currently assigning product codes to these articles through a combination of text mining and manual inspection techniques.[11]

[10] There are more than 290 software product codes (denoted as SOF) defined by CorpTech Directory. Each firm in this directory is associated with a set of self-reported product codes selected from these 290 SOF categories.

[11] Our preliminary empirical analysis reported in this draft is based on the set of OSS product introduction articles with assigned product codes.

4.2. Assigning Patents to Software Segments

To identify the relevant patents across a range of market segments from the PROMPT-CorpTech classification, our initial step was to examine the patents of specialist firms that produced in only one software segment and particularly only one six- or seven-digit CorpTech code. We thus selected 3430 single specialists from the CorpTech directory covering the 1992-2004 and 2010 years.[12] Then, we identified all patents from 1976 to 2009 held by these specialist firms, resulting in 3500 patents that can be applied to different software segments. Indeed, these patents' 3-digit US classes serve as a starting point to decide what patent classes are associated with each software segment. To fill out our concordance between patents and classes, we examine the USPTO class-subclass distribution of forward citations of our initial 3500 patents and map additional class-subclass combinations to our product market segments based on these forward citation patterns. This procedure is similar to the one used by Cockburn and MacGarvie (CM) (2009, 2011). We constructed our own classification for several reasons. First, our sample period is later than theirs so the mapping between patent technologies and product markets may have changed over time. Second, CM examined 27 specific software markets that have incomplete overlap with the open source product market segments that we are concerned with. However, in our comparison of our classification to CM's in overlapping segments, we found some similarities. Our final concordance table includes 101 software segments[13] matched to 338 patent categories (on the level of class-subclass-level1).

5. PRELIMINARY EMPIRICAL ANALYSIS AND RESULTS

5.1 Are Patents in OSS Patent Pools Less Valuable than Average Comparable Patents?

---

[12] Unfortunately, the CorpTech Directory 2005 to 2009 was not available.

[13] For convenience of analysis, most of our summary statistics will be based on 12 segments aggregated from the 101 software segments. The appendix 3 shows some examples of how the 12 aggregate segments correspond to our original 101 software segments.

In our first set of analyses we investigate some basic information about the patents in the OSS pools: the characteristics of the patents in terms of citations and claims; how the number and citations of patents in the pools are distributed across the 101 software segments and whether it is significantly different from the non-pool patents. This should provide information on the quality of patents contributed to the pool.

Table 1 provides some information on how patents in the pools compare to similar patents regarding forward citations, backward citations, and number of claims. Following the matching method employed by Jaffe, Trajtenbeg, and Henderson (1993) and followed by many others, we construct the sample of non-pool control patents by choosing the non-pool patents that belong to the same class as each of the pool patents and were granted either in 2 years before the grant year or in 2 years after the grant year of each pool patent. Of course, one concern with using forward citations during our recent sample period is that they may be truncated; as a result, we estimate the citation distribution using the structural approach described in Hall, Jaffe, and Trajtenberg (2001). As shown in the first row of table 1, pool patents' forward citations are significantly higher and backward citations significantly lower than those of control patents.[14] However, non-pool control patents have a significantly higher number of claims. Thus, while pool patents may be cited more frequently and may be less derivative than other comparable patents, they also cover a narrower technology scope than similar non-pool patents.

Table 2 further presents some evidence about how the pool patents are distributed across 12 aggregate software segments. We construct these 12 aggregate segments from our original 101 segments to facilitate the descriptive analysis. As shown by both the raw patent count and percentage, pool patents seem to be concentrated in segments that have well-established OSS

[14] We acknowledge that one other potential reason for pool patents' forward citations to be more than average patents is the effect of patent pool on improving the awareness of these patents, which make either OSS or non-OSS participants more likely to cite these patents. In our next step, we will try to address this concern.

communities such as operating systems, utility software, and DBMS. Nevertheless, when comparing this distribution with non-pool patents, another factor that drives the concentration of patents in these segments is that these segments tend to cover more categories of technologies. However, when we compared the forward citations across different segments for pool patents versus non-pool control patents, as shown in table 3, the pool patents seem to consistently have higher quality than the matched non-pool patents in most segments, though the difference between pool and non-pool backward citations is mixed.

5.2 How do Patents in the Commons Differ from Patents in OIN?

As suggested above in our discussion of the institutional difference between the Commons and OIN pools, it is important to disentangle any characteristics of the two that may have potentially different impacts on OSS entrants across different segments. Thus, we did similar comparisons as above between the Commons and OIN, presented in tables 4 through 6. First, as shown in table 4, OIN patents have significantly more forward citations, backward citations, and claims than those of the Commons. Second, table 5 shows that while the highest concentration of Commons patents is in upstream or system software, e.g. operating systems and utility software, the segments where OIN has its highest percentage of patents are more about middleware software such as communications control software, Internet tools, and DBMS.

5.3 Was IBM Contributing Less Valuable Patents to the Commons When Compared with Its Own Patent Portfolio?

Since IBM patents form such an important part of the pools in our study, we separately examine the value of patents contributed by IBM. A shown in table 7, forward citations for patents contributed by IBM to the Commons pool are significantly greater than the forward citations of

IBM's patents that belong to the same class and were granted either in 2 years before the grant year or in 2 years after the grant year of each pool patent.

5.4. How is the Variance in OSS Entry Associated with Changes in the OSS Patent Pools?

Our empirical approach is motivated by recent research that has studied how patent thickets influence market entry in the software industry (e.g., Cockburn and MacGarvie, 2009, 2011). There are several challenges for identification in our setting. First, the entry rate and the stock of pool patents may be co-determined by some segment-specific factor. For instance, for segments that are relatively more innovative and fast-moving, we would observe higher entry rate and higher rate of patenting which may lead to the increase in the number of pool patents. To address this issue, our focus is on the time series variation in the size of patent pools within a software segment and its interaction with the segment-specific density of the patent thicket. Second, one important omitted variable is that OSS incumbents may strategically contribute pool patents to certain segments which gather a number of fast-growing OSS communities providing complementary innovations. Thus, one of our next steps in our project is to identify instruments that may be correlated with the incumbents' patent contribution behavior but uncorrelated with the entry pattern over time.

Following prior literature, we use count data models with conditional fixed effects. Suppose the number of OSS entrants in software segment *j* in year *t* (denoted as Y_{jt}) follows a Poisson process and $E(Y_{jt} \mid X_{jt}, c_j) = c_j \exp(X_{jt}'\beta)$, where $X_{jt}'\beta = \beta_1 * patent\_pools_{jt-1} + \beta_2 * thickets_{jt-1} + \beta_3 * patent\_pools_{jt-1} * thickets_{jt-1} + \gamma_1 * SegmentControls_{jt} + \varepsilon_{jt}$. As a starting point, $patent\_pools_{jt-1}$ is measured by the number of pool patents related to segment *j* in year *t*-1; $thickets_{jt-1}$ is the size of patent thickets in segment *j* in year *t*-1.[15] As we are currently in the

[15] We will construct alternative proxies for the size of patent thickets within each segment in the next step.

process of constructing our measures of patent thickets, in this preliminary analysis we focus on estimating β_1. Because of the institutional difference between the two patent pools, we will first examine the overall impact of the two pools on entry and then investigate their separate impacts. That is, more specific forms on $X_{jt}'\beta$ at this stage will be:

(1) $X_{jt}'\beta = \alpha + \beta_1\, ln(1+ OSS\ pools)_{jt} + \beta_2\, ln(total\ market\ patents_{jt-1}) + \beta_3\, ln(quality\ of\ market's\ patents_{\ jt-1}) + year\_dummy_t + \varepsilon_{jt}$

(2) $X_{jt}'\beta = \alpha + \gamma_1\, ln(1+ the\ Commons)_{jt} + \gamma_2\, ln(1+ OIN)_{jt} + \beta_2\, ln(total\ market\ patents_{jt-1}) + \beta_3\, ln(quality\ of\ market's\ patents_{\ jt-1}) + year\_dummy_t + \varepsilon_{jt}$

where $ln(1+ OSS\ pools)_{jt}$ represents the log of the total patents across both pools; $ln(1+ the\ Commons)_{jt}$ represents the cumulative stock of patents in the Commons pool; $ln(1+ OIN)_{jt}$ represents the cumulative stock of patents in OIN; $ln(total\ market\ patents_{jt-1})$ indicates the cumulative stock of total patents in the product market segment; $ln(quality\ of\ market's\ patents_{jt-1})$ is used as a control for the quality of patents in the product market segment, measured by the cumulative stock of citations to *total market patents* divided by *total market patents*. The summary statistics of these variables are included in table 10. The above two specifications are preliminary: in the future we plan to add controls such as the number of incumbent firms in each market and market segment revenues. We discuss additional extensions to the model below. We also present estimates using between and within panel data OLS estimates, as well as the pooled Poisson model.

As shown in table 11 and table 12, the results show that increases in the total number of patents across both pools is associated with an increase in the number of OSS entrants in most cases. However, when we disaggregate our measure of patent pools into contributions made into the Patent Commons and OIN we find that only the Patent Commons has any effect on OSS entry.

6. FURTHER EMPIRICAL ANALYSIS

An important next step in our analysis will be to examine more carefully how omitted variable bias may be influencing our results. One particular concern for us is that unobserved product market growth may be correlated both with contributions to the pool and to OSS product introductions.

To be specific, contributions to the Commons and OIN come primarily from large, established firms. An evolving literature has begun to suggest that such firms are active in OSS in part to stimulate complementary demand for their products (e.g., Fosfuri et al. 2008). Such contributions are therefore likely to be increasing in sales growth in the complementary market. This creates an important source of potential omitted variable bias if OSS product introductions are also increasing in market segment growth, as we expect they may be.

Our next step will be to add a new instrumental variables analysis to address this concern. We plan to instrument for patent pool contributions using the percentage of contributors' sales revenue coming from each of the software product market segments in our data. These shares will come from one of two sources that we have collected data from—product market introductions as measured by PROMT (as used above) and employment by product market segment as measured by another data source, the National Establishment Time Series (NETS) database. While we have not yet incorporated information from the NETS database into our analyses, we plan to do this in our next version of the paper. We hypothesize that after controlling for the overall segment sales or employment, firms with a greater share of product market sales in a software segment may be less likely to contribute patents to the OSS community, due to a potentially larger rent dissipation effect of OS licensing. We also plan to explore a second instrument: the extent to which contributors have litigated software patents, copyrights, or trademarks in a software product market segment. This latter instrument will be a measure of the extent to which the contributing firm has attempted to monetize its IP portfolio in the software segment, and should be negatively correlated with contributions in the segment. More broadly, we plan to more carefully document empirically the incentives for patent pool contributions.

REFERENCES

Asay, M. (2010). "What Apple's and Microsoft's patent threats mean for start-ups." from http://news.cnet.com/8301-13505_3-10466670-16.html.

Bessen, J. and R. M. Hunt (2007). "An empirical look at software patents." Journal of Economics & Management Strategy 16(1): 157-189.

Cockburn, I. M. and M. J. MacGarvie (2009). "Patents, thickets and the financing of early-stage firms: Evidence from the software industry." Journal of Economics & Management Strategy 18(3): 729-773.

Cockburn, I. M. and M. J. MacGarvie (2011). "Entry and Patenting in the Software Industry." Management Science

Cohen, W. M., R. R. Nelson, et al. (2000). "Protecting their intellectual assets: Appropriability conditions and why US manufacturing firms patent (or not)." NBER Working paper.

Fosfuri, A., M. S. Giarratana, et al. (2008). "The penguin has entered the building: The commercialization of open source software products." Organization Science 19(2): 292-305.

Gilbert, R. (2004). "Antitrust for Patent Pools: A Century of Policy Evolution." Stan. Tech. L. Rev.

Graham, S. J. H. and D. C. Mowery (2005). "The Use of USPTO'Continuation'Applications in the Patenting of Software: Implications for Free and Open Source." Law & Policy 27(1): 128-151.

Hall and Helmers. (2011). "Innovation And Diffusion Of Clean/Green Technology: Can Patent Commons Help?" NBER Working paper

Hall, B. H. and M. MacGarvie (2010). "The private value of software patents." Research Policy 39(7): 994-1009.

Hall, B. H., A. B. Jaffe, et al. (2001). The NBER patent citation data file: Lessons, insights and methodological tools, NBER Working paper

Heller, M. A. and R. S. Eisenberg (1998). "Can patents deter innovation? The anticommons in biomedical research." Science 280(5364): 698.

Jaffe, A. B., M. Trajtenberg, et al. (1993). "Geographic localization of knowledge spillovers as evidenced by patent citations." The Quarterly Journal of Economics 108(3): 577.

Josh Lerner and Mark Schankerman (2010). "The Comingled Code: Open Source and Economic Development." The MIT Press

Lampe, R. and P. Moser (2009). "Do patent pools encourage innovation? Evidence from the 19th-century sewing machine industry." NBER Working paper.

Lerner, J., M. Strojwas, et al. (2007). "The design of patent pools: The determinants of licensing rules." The RAND Journal of Economics 38(3): 610-625.

Scotchmer, Suzanne (2004). *Innovation and Incentives.* Cambridge: MIT Press.

Seeker, S. (2010). "The Defensive Patent License makes patents less evil for open source."

Shapiro, C. (2000). "Navigating the Patent Thicket: Cross Licenses, Patent Pools, and Standard-Setting." Competition Policy Center, Working Paper Series.

Ziedonis, R. H. (2004). "Don't fence me in: Fragmented markets for technology and the patent acquisition strategies of firms." Management Science 50(6): 804-820.

Table 1: Pool Patents Compared to Non-pool Patents

		Pool Patents	Non-pool Control Patents	T-test
	Obs	2117	327944	
Forward citations	Mean (Std.Err.)	24.927 (0.818)	12.089 (0.040)	15.680***
Backward citations	Mean (Std.Err.)	11.121 (0.364)	16.081 (0.052)	-13.504***
Claims	Mean (Std.Err.)	20.019 (0.310)	20.897 (0.028)	-2.821***

Note: 1) Forward citations are the forward citations as of Dec 31, 2009 and are adjusted for truncations based on the HJT methods; 2) the non-pool control patents are constructed by choosing the non-pool patents that belong to the same class as each pool patent and were granted in 2 years around the grant year of each pool patent; 3) ***: significant at 1%.

Table 2: Pool Patent Count Compared to Non-pool Patent Count

Aggregate SOF Segment	Pool Patents	Percentage	Non-pool Patents	Percentage
Artificial intelligence software	32	2.84	6,602	2.36
Audio and voice processing software	38	3.37	7,950	2.84
Communications control software	121	10.74	49,710	17.75
DBMS	124	11.00	18,723	6.69
Electronic message processing and GUI software	103	9.14	13,593	4.85
Engineering and technical software	25	2.22	23,440	8.37
Financial, business practice, and management software	26	2.31	10,648	3.80
Graphics and Image processing software	89	7.90	25,090	8.96
Internet tools	97	8.61	12,757	4.56
Operating systems	180	15.97	42,730	15.26
Software development tools	84	7.45	6,972	2.49
Utility software	208	18.46	61,842	22.08

Note: 1) The above aggregate software (SOF) segments are constructed from our analysis sample's 101 software segments; 2) one patent will be multiply counted if it belongs to multiple aggregate SOF segments.

Table 3: Pool Patents Citations Compared to Non-pool Patents Citations

Forward Citations Mean Difference		Backward Citations Mean Difference	
Aggregate SOF Segment	Mean Difference (Pool - Control Patents)	Aggregate SOF Segment	Mean Difference (Pool -Control Patents)
Artificial intelligence software	5.927	Artificial intelligence software	-2.595
Audio and voice processing software	8.544**	Audio and voice processing software	25.251***
Communications control software	11.130***	Communications control software	-2.799
DBMS	16.897***	DBMS	-6.307**
Electronic message processing and GUI software	14.792***	Electronic message processing and GUI software	-8.185***
Engineering and technical software	6.261	Engineering and technical software	-0.923
Financial, business practice, and management software	11.621**	Financial, business practice, and management software	-13.475**
Graphics and Image processing software	14.901***	Graphics and Image processing software	-2.469
Internet tools	13.650***	Internet tools	-9.404***
Operating systems	8.220***	Operating systems	-1.728
Software development tools	9.860***	Software development tools	-3.796
Utility software	10.326***	Utility software	-3.767*

Note: 1) Forward citations are the forward citations as of Dec 31, 2009 and are adjusted for truncations based on the HJT methods; 2) the control patents is constructed by choosing the non-pool patents that belong to the same class as each pool patent and were granted in 2 years around the grant year of each pool patent; 3) patents in the matched sample are first collapsed to the above aggregate SOF segments and then compared with mean citations by segment; 4) *: significant at 10%; **: significant at 5%; ***: significant at 1%.

Table 4: OIN's Patents Compared to the Commons' Patents

		OIN Patents	The Commons' Control Patents	T-test
	Obs	444	1052	
Forward citations	Mean (Std.Err.)	33.955 (2.958)	17.185 (0.687)	5.522***
Backward citations	Mean (Std.Err.)	15.173 (1.590)	10.506 (0.230)	2.905***
Claims	Mean (Std.Err.)	29.756 (0.965)	18.702 (0.347)	10.775***

Note: 1) Forward citations are the forward citations as of Dec 31, 2009 and are adjusted for truncations based on the HJT methods; 2) the Commons' control patents are constructed by choosing the Commons' patents that belong to the same class as each OIN patent and were granted in 2 years around the grant year of each OIN patent; 3) ***: significant at 1%.

Table 5: OIN's Patent Count Compared to the Commons' Patent Count

Aggregate SOF Segment	OIN	Percentage	The Commons	Percentage
Artificial intelligence software	3	1.34	29	3.21
Audio and voice processing software	18	8.04	20	2.21
Communications control software	41	18.3	80	8.86
DBMS	17	7.59	107	11.85
Electronic message processing and GUI software	33	14.73	70	7.75
Engineering and technical software	1	0.45	24	2.66
Financial, business practice, and management software	7	3.13	19	2.1
Graphics and Image processing software	18	8.04	71	7.86
Internet tools	39	17.41	58	6.42
Operating systems	12	5.36	168	18.6
Software development tools	11	4.91	73	8.08
Utility software	24	10.71	184	20.38

Note: 1) The above aggregate software (SOF) segments are constructed from our analysis sample's 101 software segments; 2) one patent will be multiply counted if it belongs to multiple aggregate SOF segments.

Table 6: OIN's Patent Citations Compared to the Commons' Patent Citations

Forward Citations Mean Difference		**Backward Citations Mean Difference**	
Aggregate SOF Segment	Mean Difference (OIN –the Commons' Control Patents)	Aggregate SOF Segment	Mean Difference (OIN –the Commons' Control Patents)
Artificial intelligence software	-2.280	Artificial intelligence software	-4.476
Audio and voice processing software	-9.762	Audio and voice processing software	66.972***
Communications control software	2.342	Communications control software	5.371
DBMS	24.292**	DBMS	2.161
Electronic message processing and GUI software	22.219**	Electronic message processing and GUI software	1.683
Engineering and technical software	N/A	Engineering and technical software	N/A
Financial, business practice, and management software	19.242	Financial, business practice, and management software	-4.029
Graphics and Image processing software	29.721***	Graphics and Image processing software	1.509
Internet tools	24.271***	Internet tools	4.208
Operating systems	2.013	Operating systems	4.145
Software development tools	6.101	Software development tools	4.713
Utility software	-1.341	Utility software	2.808

Note: 1) Forward citations are the forward citations as of Dec 31, 2009 and are adjusted for truncations based on the HJT methods; 2) the Commons' control patents are constructed by choosing the Commons' patents that belong to the same class as each OIN patent and were granted in 2 years around the grant year of each OIN patent; 3) patents in the matched sample are first collapsed to the above aggregate SOF segments and then compared with mean citations by segment; 4)*: significant at 10%; **: significant at 5%; ***: significant at 1%.

Table 7: Patents in the Commons compared to IBM's other patents

		The Commons	IBM's Control Patents	T-test
Forward citations	Mean (Std.Err.)	21.908 (0.655)	20.483 (0.178)	2.099**
Backward citations	Mean (Std.Err.)	10.023 (0.180)	10.811 (0.066)	-4.106***
Claims	Mean (Std.Err.)	17.133 (0.261)	18.962 (0.090)	-6.626***

Note: 1) Forward citations are the forward citations as of Dec 31, 2009 and are adjusted for truncations based on the HJT methods; 2) IBM's control patents are constructed by choosing IBM's non-pool patents that belong to the same class as each Commons' patent and were granted in 2 years around the grant year of each Commons' patent; 3) ***: significant at 1%.

Table 8: The Commons' Patent Count Compared to IBM's Other Patent Count

Aggregate SOF Segment	The Commons' PatentS	Percentage	IBM's other patents	Percentage
Artificial intelligence software	25	2.89	429	1.90
Audio and voice processing software	20	2.31	499	2.21
Communications control software	73	8.45	2,358	10.45
DBMS	102	11.81	2,376	10.53
Electronic message processing and GUI software	67	7.75	1,733	7.68
Engineering and technical software	24	2.78	742	3.29
Financial, business practice, and management software	19	2.20	363	1.61
Graphics and Image processing software	68	7.87	1,310	5.81
Internet tools	55	6.37	1,294	5.74
Operating systems	164	18.98	4,930	21.85
Software development tools	70	8.10	956	4.24
Utility software	177	20.49	5,572	24.70

Note: 1) The above aggregate software (SOF) segments are constructed from our analysis sample's 101 software segments; 2) one patent will be multiply counted if it belongs to multiple aggregate SOF segments.

Table 9: The Commons' Patent Citations Compared to IBM's other Patent Citations

Forward Citations Mean Difference		Backward Citations Mean Difference	
Aggregate SOF Segment	Mean Difference (the Commons – IBM's Control Patents)	Aggregate SOF Segment	Mean Difference (the Commons – IBM's Control Patents)
Artificial intelligence software	3.490	Artificial intelligence software	-2.380
Audio and voice processing software	10.744*	Audio and voice processing software	0.892
Communications control software	1.895	Communications control software	0.692
DBMS	0.553	DBMS	-1.024
Electronic message processing and GUI software	0.087	Electronic message processing and GUI software	-0.288
Engineering and technical software	4.735	Engineering and technical software	0.235
Financial, business practice, and management software	4.057	Financial, business practice, and management software	-0.583
Graphics and Image processing software	2.595	Graphics and Image processing software	-0.136
Internet tools	-6.313*	Internet tools	-1.568
Operating systems	2.966	Operating systems	-0.789
Software development tools	-1.005	Software development tools	-0.857
Utility software	0.934	Utility software	-0.773

Note: 1) Forward citations are the forward citations as of Dec 31, 2009 and are adjusted for truncations based on the HJT methods; 2) IBM's control patents are constructed by choosing IBM's non-pool patents that belong to the same class as each Commons' patent and were granted in 2 years around the grant year of each Commons' patent; 3) patents in the matched sample are first collapsed to the above aggregate SOF segments and then compared with mean citations by segment; 4) *: significant at 10%; **: significant at 5%; ***: significant at 1%.

Table 10: Summary Statistics

Variable name	Definition	Observations	Mean	Std. Dev.	Min.	Max.
Segment-year-level						
OSS Entrants	The number of OSS entrants in segment j in year t	1010	0.056	0.281	0	3
OSS pools	The accumulative stock of patents in the two OSS pools associated with segment j in year t	1010	9.21	20.642	0	149
The Commons	The accumulative stock of patents in the Commons associated with segment j in year t	1010	8.282	19.522	0	139
OIN	The accumulative stock of patents in OIN associated with segment j in year t	1010	0.928	2.422	0	26
Market patents	The accumulative stock of all patents associated with segment j in year t-1	1010	2408.041	3672.653	20	29526
Quality of market's patents	The accumulative stock of citations to all patents associated with segment j in year t-1, divided by all patents associated with segment j in year t-1	1010	23.329	10.591	6.732	68.902

Table 11: Fixed-effects Regression and Between Regression

	Fixed-effects (within) regression				Between regression (regression on group means)			
Dependent variable: Number of OSS entrants	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
OSS pools	.031**				.085***			
	(.016)				(.026)			
The Commons		.034**		.041**		.084***		.096***
		(.016)		(.017)		(.023)		(.026)
OIN			-.010	-.028			.031	-.050
			(.021)	(.023)			(.049)	(.052)
ln(market patents)	-.163	-.193	-.123	-.212	-.012	-.011	.009	-.008
	(.156)	(.165)	(.145)	(.170)	(.011)	(.010)	(.010)	(.011)
ln(quality of market's patents)	-.430*	-.501*	-.415*	-.567*	.025	.047	.011	.072
	(.256)	(.278)	(.246)	(.291)	(.035)	(.035)	(.042)	(.043)
Year Dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Number of groups	101	101	101	101	101	101	101	101
Number of observations	1010	1010	1010	1010	1010	1010	1010	1010
R square	0.026	0.027	0.022	0.029	0.113	0.132	0.022	0.139

Note: 1) *: significant at 10%; **: significant at 5%; ***: significant at 1%; 2) robust standard errors clustered over segments are in parentheses.

Table 12: Poisson Regression

	Poisson				Conditional fixed-effects Poisson regression			
Dependent variable: Number of OSS entrants	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
OSS pools	.833***				.401			
	(.155)				(.281)			
The Commons		.826***		.907***	-.792	.413*		.652**
		(.152)		(.182)	(2.047)	(.281)		(.330)
OIN			.137	-.286			-.179	-.611
			(.262)	(.275)			(.355)	(.418)
ln(market patents)	-.082	-.089	.204	-.095	-.792	-.749	-1.567	-.580
	(.196)	(.197)	(.214)	(.198)	(2.047)	(2.049)	(1.914)	(2.043)
ln(quality of market's patents)	.172	.254	.207	.328	-4.941	-4.939	-7.351	-5.423
	(.531)	(.531)	(.552)	(.532)	(4.238)	(4.226)	(4.047)	(4.230)
Year Dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Number of groups					27	27	27	27
Number of observations	1010	1010	1010	1010	270	270	270	270
Log likelihood	-207.756	-206.880	-216.943	-206.283	-97.353	-97.282	-98.300	-96.205

Note: 1) *: significant at 10%; **: significant at 5%; ***: significant at 1%; 2) robust standard errors clustered over segments are in parentheses.

Appendix 1: Keywords used to search licensing software as open source in news articles

open source
open-sourced
OSS
FLOSS
source code
GPL
GNU
General Public License
Lesser General Public License
LGPL
BSD
FreeBSD
Apache License
Apache Software License
Artistic License
MIT License
Mozilla Public License
Public Domain
GPL-compatible
non-copyleft
copyleft
free software license
open source license
open-source license

Note: We noticed that it is possible for articles identified by two of the above keywords – OSS and source code – to be not related to open source product. Thus, for this subset of articles, we manually read each articles to identify whether it is related to open source product introductions.

Appendix 2: Examples of PROMPT codes

7372502	Operating systems
7372503	Operating system enhancements
7372504	Graphical user interface software
7372505	Portable document software
7372510	Software development tools
7372511	CASE software
7372512	Programming utilities
7372513	Application development software
7372514	Debugging & testing software
7372515	CD-ROM mastering software
7372520	Peripheral support software
7372521	Device driver software
7372522	Data acquisition software
7372523	Printer support software
7372530	Disk/file management software

Appendix 2: Examples of the concordance between PROMPT code and CorpTech code

Segment Name	CorpTech Code	PROMPT Product Code
Fax software	SOF-CS-F	7372650 Fax software
Database management systems (DBMS)	SOF-DM-M	7372421 DBMS
Practice management sof	SOF-HL-M	7372466 Medical practice software
Audio editing software	SOF-ME-S	7372544 Sound/audio software
Electronic bulletin board software	SOF-OA-MB	7372662 BBS software
Computer teleconferencing software	SOF-OA-MC	7372674 Videoconferencing software
Electronic message systems software	SOF-OA-ME	7372605 Electronic mail software
Groupware	SOF-OA-MG	7372630 Workgroup software
Desktop publishing software	SOF-OA-P	7372441 DTP software
Civil/struct/arch eng sof	SOF-TS-EC	7372433 Civil engineering software
Electrical/eltrnc eng sof	SOF-TS-EE	7372434 Electrical engineering software
Geographic information sys	SOF-TS-ER	7372423 Geographic information systems
Peripheral device drivers	SOF-UT-H	7372521 Device driver software
Data center mgmt software	SOF-UT-O	7372561 Data center management software
Application devel tools	SOF-UT-Q	7372513 Application development software
Data encryption software	SOF-UT-X	7372691 Data encryption software

Appendix 3: Examples of the concordance between the 12 aggregate segments and our original 101 software segments

Examples of Original 101 Software Segments	Aggregated Segment
Bridge and router software	Communications control software
Emulation/simulation sof	Communications control software
Fax software	Communications control software
Microcom-mainframe sof	Communications control software
Network configuration management software	Communications control software
Network operating systems and utilities	Communications control software
Network switching software	Communications control software
Protocol gateway software	Communications control software
Remote access software	Communications control software
Wide area network software	Communications control software
Database dictionaries	DBMS
Database management systems (DBMS)	DBMS
Database query language	DBMS
File management software	DBMS
Application devel tools	Software development tools
Code algorithms	Software development tools
Cross assembler software	Software development tools
Cross compilers	Software development tools
Debugging & testing sof	Software development tools
Development environment sof	Software development tools
Language development sof	Software development tools
Machine code linkers	Software development tools
Program translators	Software development tools
Programming utilities	Software development tools
Progrmg lang enhancements	Software development tools
System design/methodology	Software development tools
Data center mgmt software	Utility software
Data conversion software	Utility software
Data encryption software	Utility software
Disaster recovery software	Utility software
Disk/tape/file utilities	Utility software
Font utilities, letter	Utility software
Librarian utilities	Utility software
Network security software	Utility software
Performance measuring sof	Utility software
Peripheral device drivers	Utility software
Print utilities/spoolers	Utility software