Patent Pools, Thickets, and Open Source Software Entry by 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

January 2012

Abstract

We examine whether open source software (OSS) patent pools contributed by large software incumbents influence new OSS product entry by start-up software firms. In particular, we analyze the impact of a major OSS patent pool—the Patent Commons—established by IBM in 2005. We find that increases in the size of the OSS patent pool related to a software segment increase the rate of entry in the segment by startups using an OSS license. The marginal impact of the OSS patent pool on OSS entry by start-ups is increasing in the cumulativeness of innovation in the segment and the extent to which patent ownership in the segment is concentrated.

Keywords: open source, open source software (OSS), OSS entry, patent pool, patent pledge, patent thicket, start-ups' OSS innovation, cumulative innovation, concentrated patent ownership.

JEL Classification: O34, L86

1. Introduction

There has been significant growth in the development and commercialization of open source software (OSS) in recent years. A recent survey has indicated that worldwide revenue from commercializing OSS reached $1.8 billion in 2006 and is expected to reach $5.8 billion in 2011 (Broersma 2007). However, one barrier to increasing growth in OSS innovation is the risk to OSS producers of infringing existing intellectual property rights (IPR) such as patents.

While patent thickets have been shown to deter firm entry into the software industry (Cockburn and MacGarvie 2011), the nature of OSS innovation suggests that innovation in this type of software may be particularly sensitive to IPR enforcement actions for several reasons. First, the distributed, incremental development approach to developing OSS implies that innovation is highly cumulative. Moreover, this same approach makes it difficult to identify the provenance of source code, imposing high costs on OSS developers who wish to identify potentially infringing technologies. Anecdotal evidence suggests that IPR enforcement against OSS can have a significant impact on software markets, as demonstrated by the well-known set of lawsuits by the SCO Group against Linux as well as the most recent set of lawsuits among Apple, Samsung, and HTC that are related to the Android open source platform.

One way to ameliorate the patent thickets problem for OSS innovation is the creation of royalty-free patent pools, a special type of patent pool that offers royalty-free usage of patents to any firm that promises not to sue the pool's beneficiaries (Lévêque and Ménière 2007, Serafino 2007, Hall and Helmers 2011).[1] However, the empirical effectiveness of such patent pools in mitigating the patent thicket problem is largely unknown. Prior studies of patent pools have focused largely on traditional patent pools that delineate specific licensing rules and restrictions to non-pool members. These traditional patent pools have been shown to have an uncertain impact on the rate of innovation (e.g., Lerner and Tirole 2004, Lampe and Moser 2009, Joshi and Nerkar 2011).

There is also reason to question the impact of royalty-free patent pools on innovation and competition. Inventors may have insufficient incentives to provide intellectual property as a public good to the community (Gambardella and Hall 2006), and so royalty-free patent pools may contain few or low-quality patents that have little impact on innovative activity. One software industry commentator, for instance, suggests that "the perception is that bigger companies only commit their least-effective, least-important patents to a patent pool" (Seeker 2010). In their recent analysis of patents pledged to the "Eco-Patent Commons," established in 2008 to provide royalty-free access to third parties to patented technologies related to climate change, Hall and Helmers (2011) find mixed evidence on their impact on

---

[1] Hall and Helmers (2011) refer to a royalty-free patent pool as a patent commons. We have found both labels are currently in use, and use the label royalty-free patent pool both to draw a distinction to traditional patent pools and to avoid confusion with the OSS "Patent Commons" that was developed by IBM and is a focus of this study.

innovation in a different industry setting. While their analysis is preliminary—due to the short period of time following the establishment of this particular patent pool—the authors conclude that patents contributed to the Eco-Patent Commons have "no discernible impact on the diffusion of the knowledge embedded in the protected technologies to other patenting firms" (Hall and Helmers 2011).

Motivated by these observations, we take a first step to evaluate how innovation is influenced by the creation of a royalty-free patent pool that is made available to the OSS community. Specifically, we examine the impact of such an OSS patent pool on entry by start-ups using an OSS license (which we refer to as "OSS entry").[2] Our focus on this margin of innovation is motivated by several considerations. First, the formation of OSS patent pools may have a particularly strong impact on start-up innovation. Unlike large firms, start-ups usually lack the R&D capabilities and financial resources required to expand their own patent portfolios, so it is difficult for them to navigate patent thickets using other approaches such as cross-licensing agreements.[3] This is particularly likely to be the case for start-ups that produce OSS and who, for a variety of reasons, may be unlikely to patent their innovations. Further, as has been highlighted elsewhere, our knowledge of the implications of formal IP rights for OSS innovation is still quite limited (Lerner and Tirole 2005a, von Hippel and von Krogh 2003). Our focus on entry is motivated by our context: since many firms producing under an OSS license do not patent, traditional patent-based measures of innovation are inappropriate for our setting. In short, due to the uncertainty about the effectiveness of OSS patent pools, our research strategy is to examine their impact along a margin of innovation where they are more likely to matter. We leave the implications for other types of innovation (such as entry under proprietary licenses and the behavior of large firms) for future work.

To motivate our empirical analysis, we develop a model that shows how an OSS patent pool can change the bargaining game between a proprietary incumbent patent holder and a start-up firm. The model, which builds on prior work by Llobet (2003) and Galasso and Schankerman (2010), shows that increases in the size of the pool influence the outcome of the litigation game and consequently the start-up's OSS entry decision. Comparative statics from the model show that (i) changes in the size of the OSS patent pool related to a software segment facilitates OSS entry by start-up firms into the same segment; and (ii) the marginal effect of the pool on OSS entry will be especially large in software segments where the cumulativeness of innovation is high or where patent ownership in a segment is concentrated.

Focusing on one major OSS patent pool—the Patent Commons—we examine the empirical salience of these predictions using a unique data set. We assemble data on OSS entry using data on

---

[2] We similarly refer to firms that have engaged in OSS as "OSS firms." More precisely, an OSS firm is defined as one that develops, uses, or commercializes source code that meets the Open Source Initiative definition of open source software.

[3] As noted by Matt Asay, the chief operating officer at Canonical (the company behind the Ubuntu Linux operating system), "this [type of patent collective] may be the only refuge for start-ups and others, like Red Hat, that don't have an aggressive patent-acquisition policy." (Matt Asay 2010)

product releases from 2,054 start-up software firms contained in the Gale database "PROMT". Following prior work that has examined the extent to which patents deter entry into the software industry (Cockburn and MacGarvie 2009, 2011), we allocate patents to software product market segments by (i) identifying the main technological classes of patents acquired and cited by single software product producers and (ii) comparing a set of keywords from a software segment classification with the keywords from the patent's technological classes. We use this mapping to identify the number of pool patents related to each market segment, as well as the cumulativeness of innovation and patent ownership concentration in the segment.

Using count data conditional fixed effects models, our empirical strategy examines whether time series variation in the number of patent claims in the OSS patent pool related to a software segment is associated with changes in the number of OSS entrants into that segment. Our initial approach is to assume that changes in the number of pool claims are uncorrelated with omitted variables that may influence OSS entry by start-ups. We later relax this assumption, employing count data models with instrumental variables using Generalized Method of Moments (GMM) estimation. Our first instrument uses the pre-sampling stock of patents held by the major patent pool contributor—IBM—across each of the market segments in our sample. We interact this variable with time dummies to obtain time series variation. Our second instrument uses the number of IBM patents that were opposed at the European Patent Office in each software segment-year.

Our results suggest that a 10% increase in the pool's patent claims in a software segment is associated with a 1.5%-2.9% increase in the rate of OSS entry by start-ups into that segment.[4] The marginal impact of the OSS patent pool is significantly greater in segments where the cumulativeness of innovation is high: a 10% increase in the pool's patent claims is associated with a 3.8%-5.6% increase in the rate of OSS entry when the cumulativeness of innovation is at its 90th percentile, compared to no significant increase when cumulativeness is at its 10th percentile. The effect of OSS patents pools is also greater when the concentration of patent ownership is high. A 10% increase in the pool's patent claims is associated with a 1.3%-1.7% increase when patent concentration is at its 90th percentile and no increase when patent concentration is at its 10th percentile, however the statistical significance of these results vary somewhat across specifications. We explore the robustness of all of our results to adding a variety of controls, as well as to GMM instrumental variables estimation. These additional analyses suggest a causal interpretation to our results.

Our study provides the first large sample evidence on how the provision of a royalty-free patent pool shapes OSS entry by start-ups. As such, our research adds to the literature that looks at how IPR licensing and enforcement influences OSS innovation (Graham and Mowery 2005, Lerner and Tirole

---

[4] All marginal effects are evaluated at mean values of covariates.

2005b, Maurer and Scotchmer 2006) as well as recent work that studies firm decisions to commercialize innovations using an OSS license (Bonaccorsi et al. 2006; Dahlander 2007, Fosfuri et al. 2008).

Our research also adds to recent studies that empirically investigate the economic implications of patent pools. While previous studies on patent pools have focused mainly on their design (e.g., Lerner and Tirole 2004, Lerner et al. 2007, Layne-Farrar and Lerner 2010), a few recent empirical studies (Lampe and Moser 2010, Joshi and Nerkar 2011, Hall and Helmers 2011) have begun to look at the impact of patent pools on the direction of innovative activities. However, to the best of our knowledge, there has been no empirical research on how patent pools shape start-up entry.

Finally, we also contribute to the literature by examining the potential anti-commons problems from strategic patenting and the impact of patent thickets on entry into the software industry (e.g., Cockburn and MacGarvie 2011, Ziedonis 2004). While the patent thickets problem can be examined from different perspectives, we highlight the roles of cumulativeness of innovation and patent ownership concentration as two different and important dimensions of patent thickets. We propose mechanisms under which these characteristics may interact with the OSS patent pool to determine start-up entry costs. Thus, our research also provides empirical evidence on the effectiveness of mechanisms meant to mitigate the anti-commons problem, such as the establishment of patent pools or standard-setting organizations (Shapiro 2001, Rysman and Simcoe 2008).

2. Theory and Hypotheses

We study the implications of a royalty-free patent pool for entry into an OSS market segment. While we describe the institutional characteristics of these patent pools below, their key features are that licensees are neither required to pay a license fee nor are they required to contribute patents to the pool. Patent pool licensees can use the patent pool defensively when facing litigation from non-pool patent holders. In this section we develop a simple model that delineates how the size of an OSS patent pool, together with the nature of innovative activity and the distribution of patent holdings in the segment, shapes start-up decisions to enter into that segment by releasing a new product under an open source license. This model builds upon and extends recent work by Llobet (2003) and Galasso and Schankerman (2010).

Consider a start-up's decision to enter into a software segment by introducing a new OSS product that could generate a revenue V.[5] Suppose this market segment has a set of patented technologies held by n different parties that the start-up has to navigate through to successfully enter into the segment. We assume these parties are symmetrical in importance. We examine the case in which the start-up already has access to n-1 of the required patents that would enable it to generate a value $V' < V$, and study how

---

[5] While a firm may produce both open and closed source software (e.g., Bonaccorsi et al. 2006, Lerner and Schankerman 2010), it often does not view these two as substitutes when considering whether to enter a market segment. This viewpoint is consistent with empirical evidence from our data.

the existence of the patent pool affects the *n*th negotiation. The start-up's OSS product uses some code developed by the open source community substituting the *nth* party's patented technology to make an improvement Δ on the code.[6] In other words, product improvements are ordered as a quality ladder, as in O'Donoghue et al. (1998) and Scotchmer (2005). If it enters, the start-up also faces an irreversible investment *c* associated with developing and commercializing the new product. Upon the start-up's introduction of this new OSS product, the *n*th party can make a settlement offer to the start-up.

Our analysis proceeds in two stages. First, we show how the size of the OSS patent pool related to a segment influences the litigation game outcome and consequently the start-up's entry decision. Second, we examine how the marginal effect of the pool is influenced by the value at stake in the *n*th negotiation, which is in turn determined by the features of the patent thickets in the market segment. In particular, we focus on two features of patent thickets stressed by prior literature (Noel and Schankerman 2006, Cockburn and MacGarvie 2011). The first is the cumulativeness of innovation within a software segment, defined as the extent to which the innovations within a segment are related with or build upon each other. Second, we study the concentration of patent ownership within a software segment. In our setting, the concentration of patent ownership translates to the number of parties holding the set of essential blocking patents with whom the start-up must negotiate.

Following Galasso and Schankerman (2010) and Lerner and Tirole (2004), we use the following equation to capture the value at stake for the *n*th negotiation.

$$v(n, \theta, V) \equiv V - V' = V - V \frac{n-1}{n} \theta \tag{1}$$

The value at stake in the *n*th negotiation can be written as the difference between *V* and *V'*. We use $V' = V \frac{n-1}{n} \theta$ to indicate the value obtainable using only the *n*-1 patents that are already held by the start-up. Equation (1) shows how characteristics of the patent thicket influence the value of the *n*th negotiation. First, *n* is the total number of parties that hold blocking patents. A decrease of *n* (an increase of concentration of patent ownership) increases the value of the *n*th round of negotiation *v*, i.e., $\partial v / \partial n < 0$. The parameter θ captures the extent of cumulativeness in innovation and will influence the value of the *n*th negotiation: For example, when innovation in a market segment is very cumulative (lower θ), the value at stake in the *n*th negotiation, *v*, is very high: it is equal to *V* if $\theta = 0$. In contrast, when innovation in a market segment is non-cumulative (high θ), the patents held by each of the *n* parties will independently contribute value to the new product. In the limit, when $\theta = 1$ the value at stake in the *n*th negotiation will be equal to $\frac{1}{n} V$. Therefore, $\partial v / \partial \theta < 0$.

---

[6] Alternatively, the code could be developed internally and issued under an OSS license. While the following discussion still holds under this case, this would add an additional strategic decision for the start-up firm: whether to enter under an open source or proprietary license.

Our litigation game is motivated by the model introduced by Llobet (2003). The structure and timing is shown below.
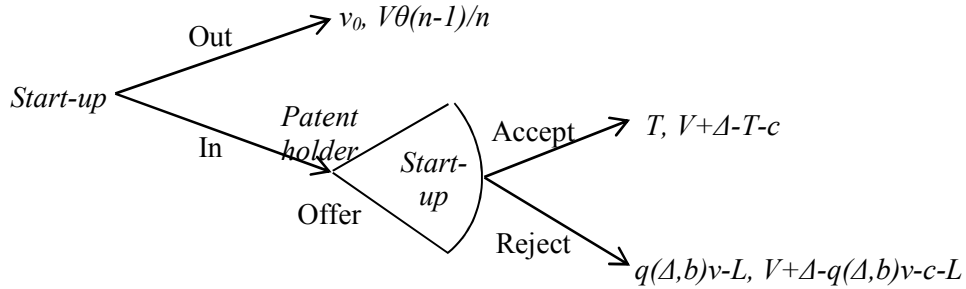
In the first stage, the start-up decides whether to enter into a software segment by introducing a new OSS product. If the start-up decides not to enter, given its current access to n-1 patents, its payoff will be $V' = V\frac{n-1}{n}\theta$;[7] the nth party will enjoy the amount of v_0 from other potential licensees. If the start-up decides to enter, the patent holder will make a settlement offer to the start-up with a license payment T. Following Llobet (2003), suppose the size of the improvement Δ is private information and unobservable to the patent holder; for simplicity, suppose the patent holder knows that Δ is drawn from a uniform distribution $G(\Delta)= \Delta$ with the density function $g(\Delta)=1$. The patent holder chooses T to optimize its expected payoff based on the distribution of Δ. Therefore, if the start-up accepts the offer, the payoffs will be:

$$\pi_s^S(\Delta, v, T) = V + \Delta - T - c \tag{2}$$

$$\pi_p^S(\Delta, v, T) = T \tag{3}$$

for the start-up and patent holder, respectively.

If the start-up firm rejects the offer, the patent holder can commit to litigate. Regardless of the outcome of the litigation, both parties incur a litigation cost denoted as L_p for the patent holder and L_s for the start-up. For simplicity, we assume $L_p=L_s=L$. During the legal proceedings, the courts receive a signal on the true value of the patents and decide whether the start-up's invention is infringing or not based on two factors: (i) how much the invention improves upon the infringed patent (Δ) and (ii) the validity of the patent holder's patents b. For simplicity, b is assumed as $b \in [0,1]$: $b=1$ implies the patents are valid while conversely $b=0$ implies the patents are invalid. The probability that the courts will favor the patent holder, $q (\Delta, b)$, is modeled as the following form for simplicity: $q(\Delta, b) = b(1-\Delta)$.[8]

---

[7] This value may be realized through a variety of different approaches, such as re-licensing the n-1 parties' patents or attracting venture capital funding.

[8] That is, when $b=1$ and $\Delta=0$, the courts will rule that the start-up is liable and be made to compensate the patent holder. When the start-up makes a significant improvement on the patented technologies, e.g. $\Delta=1$, even if $b=1$, the courts will rule that the start-up is not liable. On the other hand, when the patent is not valid, i.e. $b=0$, no matter how much improvement the start-up makes (for any value of Δ), the start-up will not be liable.

We argue that OSS patent pools can be used to strengthen a start-up's negotiating position when bargaining over blocking patents that are not part of the pool. Because the patents in the pool can be used as prior art during litigation, pool licensees can leverage the pool's patents to invalidate the infringed patents.[9] Such prior art searches have been used to invalidate infringed patents in several well-known cases where OSS was alleged to be infringing, in particular the cases of *Firestar/Datatern v. Red Hat* (Dillon 2008, Paul 2008a) and *Trend Micro v. Barracuda Networks* (Paul 2008b). The greater the number of patents that have been contributed to the pool, the more likely it is that courts will identify prior art from the pool to challenge the nth party's patents.[10] Thus, we expect b to decrease in the size of x, $\partial b/\partial x < 0$.

If the start-up is found liable, the court awards the patent holder payment v from the start-up. This v represents the payment under the unjust enrichment doctrine, the amount that the start-up would earn from successfully infringing this patent, given that it had access to the other n-1 proprietary incumbent patents (Galasso and Schankerman 2010, Schankerman and Scotchmer 2001). Thus, the payoff for the patent holder if it wins the lawsuit should be v-L and for the start-up should be $V+\Delta$-c-L-v. If the start-up is found not liable, the patent holder will lose the royalty payment v for the patent[11] and incur a payoff -L; the start-up will receive payoff $V+\Delta$-c-L. Therefore, the expected benefits to the patent holder and the start-up if the start-up rejects the settlement offer can be expressed as follows:

$$\pi_s^l(\Delta, v, b) = V + \Delta - q(\Delta, b)v - c - L \tag{4}$$

$$\pi_p^l(\Delta, v, b) = q(\Delta, b)v - L \tag{5}$$

We denote the strategy of the start-up as σ_s, which includes two decisions: whether to enter by introducing a new OSS product and whether to settle with the patent holder. First, there exists an improvement threshold $\underline{\Delta}$ that makes the start-up indifferent between not entering (with a payoff $V\theta\frac{n-1}{n}$) and entering. This $\underline{\Delta}$ can be defined from the equation (6) below. Second, there is a threshold of $\Delta^s(T)$ that makes the start-up indifferent between settlement and going to court: a start-up with improvement larger than $\Delta^s(T)$ will be more confident that it can win at trial and will therefore choose to go to court; for Δ smaller than $\Delta^s(T)$ but larger than $\underline{\Delta}$, the start-up would enter but rather settle. From equations (2) and (4), we can implicitly define $\Delta^s(T)$ using equation (7).

---

[9] To fulfill this defensive role, the Patent Commons project (the major OSS patent pool within the OSS community) established a partnership with the USPTO to ensure that patent examiners have access to all available prior art in the pool relating to the patents in question. See http://www.patent-commons.org/news/index.php?displaynews=17&page=1 for more details.

[10] Increases in the size of the patent pool will also strengthen the start-up's negotiating position if it enables the latter to sue the incumbent firm for infringement of one or more of the pool patents. For example, in the *Trend Micro v. Barracuda Networks* case, Barracuda countersued Trend Micro using a portfolio of antivirus patents obtained from IBM (Paul 2008b). As we show below, the pool could also effectively reduce the threshold for OSS entry through countersuits.

[11] Note that if the patent holder loses, it may not be able to collect the licensing fee v from any potential licensee. This is because this ruling decision will make it attractive for the potential licensee to adopt the free open source code with functions similar to the patents held by the nth party but with less cost.

$$V + \underline{\Delta} - T - c = V\theta \frac{n-1}{n} \tag{6}$$

$$\mathrm{T} = q(\Delta^s, b)v + L = b(1 - \Delta^s)v + L \tag{7}$$

Therefore, the strategy of the start-up as σ_s can be expressed as follows.

$$\sigma_s \begin{cases} = \text{(out) if } \Delta <= \underline{\Delta} \text{ (i.e., do not enter)} \\ = \text{(in, accept the offer) if } \underline{\Delta} <= \Delta <= \Delta^s(\mathrm{T}) \text{ (i.e., enter and pay amount T for licensing fee)} \\ = \text{(in, reject the offer) if } \Delta >= \Delta^s(\mathrm{T}) \text{ (i.e., enter and go to trial)} \end{cases}$$

As assumed by Llobet (2003), while the patent holder does not know the size of the improvement Δ, it holds a belief $\widetilde{\underline{\Delta}}$, above which the start-up will enter. Conditional on this belief, the patent holder will decide T that maximizes its expected payoff $\pi(v, b, L | \widetilde{\underline{\Delta}})$:

$$\max_T \pi(v, b, L | \widetilde{\underline{\Delta}}) = \max_T \int_{\Delta^s(T)}^1 (q(\Delta, b)v - L) \frac{g(\Delta)}{1 - G(\widetilde{\underline{\Delta}})} d\Delta + T \cdot \frac{G(\Delta^s(T)) - G(\widetilde{\underline{\Delta}})}{1 - G(\widetilde{\underline{\Delta}})} \tag{8}$$

We follow Llobet (2003) to define a pure strategy sequential equilibrium of the litigation game to be a strategy profile (T^*, σ_s^*) such that (i) T^* maximizes the expected profits for the patent holder in equation (8) given a belief $\widetilde{\underline{\Delta}}$ and (ii) for all Δ, $\sigma_s^*(\Delta)$ maximizes the expected profit for the start-up given T^*.

Lemma 1: The unique pure strategy sequential equilibrium $\underline{\Delta}^*$ (the optimal improvement threshold for OSS entry) has the following form[12]:

$$\underline{\Delta}^* = (1 + vb)^{-1}(vb + c - L - v) \tag{9}$$

Proposition 1: The size of the patent pool x will decrease the minimum improvement threshold for OSS entry $\underline{\Delta}^*$. That is, $\frac{\partial \underline{\Delta}^*}{\partial x} < 0$.

The above proposition highlights the effect of an OSS patent pool on OSS entry. By decreasing the likelihood that the courts will uphold the patent,[13] the OSS patent pool will reduce the expected license payment offered by the patent holder. This decreases the threshold of improvement for OSS entry, meaning that start-ups with smaller improvements (Δ) will enter.

We now consider how the efficacy of the patent pool is affected by two features of patent thickets—the cumulativeness of innovation and the concentration of patent ownership. We argue that the key mechanism through which these characteristics influence entry lies in how they shape the incremental value obtained from the focal nth negotiation (i.e., the size of v). Therefore, we first introduce the

---

[12] We provide proofs for this lemma and for all other lemma and propositions in the Appendix A.

[13] If we assume the incumbent's patents to be valid and consider the countersuing effect of the patent pool instead, the start-up will be liable with probability $1-\Delta$. However, if the start-up uses the pool patents to countersue, one can interpret the parameter b as a discount factor for the damage payments v imposed by the court. That is, the start-up will face the reduced damage payments bv and the expected payment from the start-up to the patent holder will be $(1-\Delta)\,bv$. For example, b could represent the effect of ex post negotiation between incumbent and start-up that reduces the size of the damages payments to the incumbent. Mathematically, the model is unchanged and the equilibrium threshold for OSS entry under this alternative interpretation is the same as our baseline model where we model the effects of the patent pool on the infringed patents validity.

following lemma related to the interaction of an OSS patent pool and the value at stake for the nth negotiation:

Lemma 2: The threshold for OSS entry $\underline{\Delta}^*$ will be lower when both the size of the patent pool x is large and the v is high. That is, $\frac{\partial^2 \underline{\Delta}^*}{\partial x \partial v} < 0$.

Lemma 2 suggests that the marginal effect of the OSS patent pool in lowering the licensing payment T will be most important when v is high. Intuitively, consider the start-up's entry decision, which depends on both the value (i.e., $\Delta + v$) it can obtain from the focal technologies and the potential costs (i.e., $T + c$). When the size of the patent pool is relatively large, it will reduce the bargaining position of the nth party, which translates into a lower licensing payment T for the start-up; further, if the start-up can obtain very high value from the focal technologies, the start-up will become more sensitive to the benefits of the OSS patent pool.[14] These two parameters (x and v) will interact to reduce the improvement threshold for OSS entry.

From equation (1), we know that when the cumulativeness of innovation is high (a low θ), the surplus the start-up can obtain from embedding the focal technologies will be high (a high v). Thus, the effect of the OSS patent pool will be stronger when both θ is low (a high v) and x is large. Similarly, a higher patent ownership concentration (lower n) will also lead to a higher value of the focal technologies (a high v) and thus interact with the effects of the OSS patent pool on the start-up's entry decision. As formally proved in the Appendix A, we have the following propositions.

Proposition 2: The threshold for OSS entry $\underline{\Delta}^*$ will be lower when both the size of the patent pool x is large and the cumulativeness of innovation is high.

Proposition 3: The threshold for OSS entry $\underline{\Delta}^*$ will be lower when both the size of the patent pool x is large and the concentration of patent ownership is high.

3. Research Setting

We define an OSS patent pool as a collection of patents pledged to OSS firms for royalty-free usage. We focus on one major OSS patent pool—the OSDL's Patent Commons project (denoted as "the Patent Commons" hereafter). In January 2005, IBM pledged 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." Subsequent to IBM's action, several other incumbents that participate in OSS software pledged around 30 patents to this pool.[15] "Pledge" in this

---

[14] Conversely, any factor that reduces the value of the nth negotiation will also reduce the effect of the OSS patent pool on entry. For instance, it can be shown that if transaction costs associated with each negotiation (assumed to be zero in the model) are high, the value at stake in the nth negotiation is lower, and the effect of the patent pool will be reduced. This also suggests that our Proposition 1 is entirely driven by the effect of the OSS patent pool on the bargaining power of the startup.
[15] Example companies include Computer Associates International Inc. and Open Invention Network, LLC.

context means to offer patents royalty-free to any third party that (i) is engaging in activities that might otherwise give rise to a claim of patent infringement and that (ii) promises not to sue the pool's beneficiaries (Patent Commons project 2005).[16] IBM announced in its press release that it believed this was the largest patent pledge of any kind. The introduction of this pool was expected to confer several benefits upon producers of OSS. Participants in the Patent Commons can freely embed technologies covered by these patents into their own products without any fear of being sued. Further, all pledged patents are explicitly listed on an online public database, and there is no need to sign any formal agreement between the Patent Commons and the beneficiary of the pool to use them.

It is worth noting that, while we found a series of recent patent pledging events based on a search of major news outlets (see Appendix B for a detailed summary of these events), our choice of the Patent Commons as the focus of our analysis is guided by several factors. First, the patent pool must be economically important in the sense that it comprises a large collection of patents. Second, since we are interested in measuring how the effect of OSS patent pools vary across different software segments, the patent pool must cover multiple software technology markets. Third, because we expect some time lag between the announcement of the royalty-free patent pool and entry by software start-ups, the time window between the patent pool contribution and the end of the sampling period should be long enough to observe any significant behavioral changes. Fourth, the patent pool needs to specify the contributed patents on a very detailed level, by, for example, listing the available patent number. The Patent Commons is the only OSS patent pool that met all four of these criteria. Nevertheless, in our empirical analysis we control for the effects of other patent pool-like institutions.

4. Data

4.1 Sample

Our sample consists of 2,054 US software firms from the 2004 and 2010 editions of the CorpTech Directory of Technology Firms[17] (denoted as CorpTech 2004/2010 hereafter) that primarily operate in the prepackaged software industry. We combine this sample with data from the National Establishment Time Series (NETS) Database, which includes 100,000 U.S.-based firms with primary SIC 7372. Our use of two data sources reflects constraints with each. The CorpTech data have detailed information on the product market segments of firms, but have little time variation, while the NETS data have limited product market information but vary over time.

---

[16] For more details, see http://www.patent-commons.org/resources/about_commitments.php.
[17] 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. The combination of CorpTech 2004 and 2010 is to address potential survivor bias.

As noted above, the focus of our study is on start-up firms. As a result, we restrict our sample to firms that were founded after 1990 and that have fewer than 1000 employees and less than $500 million in annual sales.[18] Our sampling period is from 1999 to 2009, with 6 years before the establishment of the Patent Commons and 5 years after. We believe this time window is sufficiently long to capture the impact of the Patent Commons on OSS entry.

4.2 Identifying Software Segments and the Matching Patent Classes

We use the product code classification system embedded in the Gale database "PROMT" (Fosfuri et al. 2008) as our primary source to define software market segments.[19] Because of two major drawbacks of relying only on the PROMT classifications (we describe these in further detail in the Appendix C), we further consolidate PROMT product categories with CorpTech's "SOF" product classes[20] to create a PROMT-CorpTech concordance. Under this concordance, each PROMT product code is associated with a detailed set of keywords. The keywords for each product class are used to (i) manually assign PROMT product codes to PROMT news articles with missing codes and (ii) match software segments with the most relevant patent classes as described below.

An important part of our data construction involves matching product market segments to patents. This allows us to identify both the cumulativeness of innovation and the concentration of patent ownership in a software segment. As is well-established in the literature, this matching is a challenging undertaking (e.g., Griliches, 1990, Silverman 1999). To facilitate our mapping between software products and patents, we follow Cockburn and MacGarvie (2006, 2011) and match software patents to CorpTech "SOF" product classes to create a patent-CorpTech concordance. Because our software segments are classified through PROMT categories, in order to create the final mapping between software segments and patent classes, we then combine the PROMT-CorpTech concordance and patent-CorpTech concordance to form the PROMT-patent concordance. The final concordance that we use in the empirical analysis consists of 33 software segments matched to 422 patent class-subclass combinations[21] (see the Appendix C for a detailed discussion of our data construction process).

5. Measures

5.1. Dependent Variable: OSS entry

---

[18] Our results are robust to the use of alternative thresholds for inclusion in our sample. For example, our results are robust to an alternative sample of start-ups that includes firms founded after 1990 that have fewer than 500 employees and less than $500 million annual sales.

[19] A few examples of PROMPT product codes are included in Table A-1 in Appendix C.

[20] 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.

[21] Table A-3 in Appendix C lists examples of this final concordance between software segments and US patent class-subclass combinations.

This variable refers to the number of OSS entrants into software segment *j* in year *t*. It is to capture the threshold to entry ($\underline{\Delta}^*$) in our propositions. We use a three-step procedure to identify new OSS entry in a software segment based on the press releases of the 2,054 firms in the PROMT database. First, following work by Fosfuri et al. (2008) and Bessen and Hunt (2007), we searched for a set of keywords within PROMT articles to identify articles related to OSS. Appendix C includes the full set of keywords. Second, we *manually* read all search results that included words from the first step to identify new OSS product introductions. We identified an article as referring to an introduction of a new OSS product when the article indicated that either of the following took place: (i) the introduction of a new software product that offered one or more of its module(s)[22] under an open source license (we label such modules as *open source modules*); and (ii) the introduction of a new version of an existing software product with open source modules. Third, to identify entry we kept only the first open source module release by a start-up into a segment. In total, we have 242 new OSS product entries made by 85 start-up firms from 1999 to 2009.[23] We aggregated these new OSS product entries by software segment and year. Our dependent variable is therefore equal to the number of new OSS start-up entrants in segment *j* and year *t*. The data are structured as a balanced panel. Table 1 includes a brief description of measures and summary statistics for all variables used in our empirical analysis.

5.2. Independent Variables

OSS patent pool. This variable refers to the number of patents that belong to the Patent Commons that are related to software segment *j* in year *t*. It corresponds to the OSS patent pool size *x* in the model. As discussed by Merges and Nelson (1990), it is the scope of a patent that determines the patent's economic and legal significance. In a setting with cumulative technologies, broader pool patents will be more likely to invalidate blocking patents. To capture these effects, we measure the claims-weighted count of patents in the Patent Commons pool related to each software segment.[24] We further take the logged value of this variable[25] to reduce skewness.

Cumulativeness. This variable refers to the cumulativeness of innovation within segment *j* in year *t* and is negatively correlated with the parameter *θ* in the model. To measure this concept we use patents' backward citations, which provide information about "existing ideas used in the creation of new ideas" (Caballero and Jaffe 1993) and indicate "some form of cumulative technological impact" (Jaffe et al. 1998). Following Clarkson (2005), we measure it based on the average propensity for patents in segment *j*

---

[22] In software, a module is a part of a program. A software product is composed of one or more modules that are linked together but preform different functions (e.g. the calendar module available in the Microsoft Office's Outlook).

[23] This procedure implicitly assumes there is no OSS entry by firms prior to 1999. We believe this assumption is supported by empirical evidence. For example, SourceForge, a major repository of OSS, was started in November 1999.

[24] We also use raw patent counts as a robustness check. The results are qualitatively similar to this claims-weighted measure.

[25] We add 1 to the variable when taking the log.

and year t to backward cite patents within the same segment j. This is roughly similar to the way economists have measured the cumulative nature of innovation at the firm level, e.g. using the extent to which firms self-cite their own patents (Hall et al. 2005). In our setting, we proceed as follows. If we sort the N patents within a software segment j chronologically (with $m=1$ being the oldest patent and $m=N$ being the youngest), the cumulativeness for each patent n (i.e. the propensity for patent n to cite preceding patents within the same segment) is calculated as $C_n = \sum_{m=1}^{N} \frac{x_{nm}}{n-1}$, where x_{nm} is a dummy variable equal to one if patent n back-cites patent m, and zero otherwise (with both patents belonging to the same segment), $(n\text{-}1)$ is the total number of possible citations, and $n>1$, since C_1 is undefined. In other words, the cumulativeness of a focal patent in segment j is based on the share of potential backward citations to patents belonging to the same segment that are actually cited by the focal patent. The cumulativeness of innovation for software segment j is then the average of all N-1 patents' cumulativeness:[26] $C_j = \frac{\sum_{n=2}^{N} \sum_{m=1}^{N} \frac{x_{nm}}{n-1}}{N-1}$. This measure varies over time based on the grant year of the segment j patents under consideration. Notice that the oldest patents in a segment tend to have greater cumulativeness since the potential number of patents that can be cited is smaller. As a robustness check, we also used an alternative weighting scheme, one that provides relatively lower importance to the cumulativeness measure of older patents. As in Clarkson (2005), it is calculated as $C_j = \frac{\sum_{n=1}^{N} \sum_{m=1}^{N} x_{nm}}{N(N-1)/2}$. For both measures, we take the logged value to reduce skewness.

Concentration. This variable indicates the extent of concentration of patent ownership in a segment and is negatively correlated with the parameter n in the model. Following Noel and Schankerman (2006) and Cockburn and MacGarvie (2011), we use the four-assignee citation concentration ratio to measure the concentration of patent ownership in a software segment. Backward citations indicate the extent to which a technological area has already been covered by prior art, so the share of backward citations owned by an assignee suggests the extent to which the assignee holds existing patented technologies and therefore the importance of negotiating with the assignee. To construct this variable, we first calculate the number of citations made by patents in segment j up to year t that are held by the cited assignee n (denoted as s_{njt}). Then we arranged s_{njt} in descending order. The total citations owned by the four firms that received the top four largest number of citations made by patents in segment j in year t (i.e. the top four s_{njt}, where $n=1,2,3,4$) is $\sum_{n=1}^{4} s_{njt}$. Thus, the four-assignee citation concentration ratio for

---

[26] The average only considers N-1 patents since C_1 is undefined.

segment j in year t is calculated as $\frac{\sum_{n=1}^{4} s_{njt}}{total\_citations_{jt}}$, where *total_citations*$_{jt}$ is the total number of citations made by patents in segment j up to year t.[27]

5.3. Control Variables

Sales. One important factor that may correlate with both the behavior of firms contributing to the Patent Commons and OSS entry by start-ups is the size of the market in software segment j, which is proxied by the total sales in segment j in year t. Because we do not have CorpTech data between 2005 and 2009, we use NETS data to measure this variable. Roughly 4,500 software firms in the NETS data are assigned to one of the eight-digit SIC categories (e.g., 73729901) that correspond to eight broad categories in the software industry. We compute the total sales for each of the eight SIC categories and then map them to our 33 software segments and use the matched sales to approximate the overall sales for each segment.

 Potential licensors. The costs associated with patent thickets are determined not only by the concentration of patent ownership in a segment, but also by the total number of different parties holding patents essential to a segment. We compute the total number of assignees who hold citations made by all patents in segment j up to year t to measure the total potential licensors with whom a start-up would have to negotiate. We further use the logged value to reduce the skewness of this variable.

 Total patents. Although we are most interested in two of the most important features of patent thickets—the cumulativeness of innovation and the concentration of patent ownership, the total number of patents related to a market has also been used as a measure of the density of patent thickets (Cockburn and MacGarvie 2011). We add this variable as an additional control to isolate this effect, computing the claims-weighted patent count related to each software segment j cumulated up to year t.

 Patent quality. This variable is a control for the quality of patents in the market segment j in year t. As has been noted elsewhere, higher quality patents suggest superior technological capabilities possessed by existing incumbents in the segment, which leaves less room for start-ups to innovate further. This variable is equal to the log value of the cumulative stock of citations received by the patents in segment j (adjusted for truncations) divided by total number patents in j up to year t.

 Open Innovation Network (OIN) patents. At the end of our sample period, another OSS patent pool institution—OIN—was established. Similar to the Patent Commons, OIN offers contractually royalty-free usage of these patents to OSS participants as long as users promise not to file suit against software associated with the Linux System.[28] We do not focus on this pool in our main analysis as it was introduced too late in our sample period to have a measureable effect on entry during our sample.[29]

---

[27] We also use an eight-assignee citation concentration ratio as a robustness check. The results are qualitatively similar to this four-assignee citation concentration ratio measure.

[28] For the detailed definition of the Linux system, see http://www.openinventionnetwork.com/pat_linuxdef.php.

[29] For the 130 patents contributed to OIN from year 2006 to 2009, 70 percent were contributed in 2008 and 2009.

However, we include it as a control. We measure this variable as the claims-weighted patent count of OIN patents related to software segment j cumulated up to year t.

   Standard-setting organization (SSO) patents. As mentioned earlier, another important mechanism to address the anti-commons problem is SSOs. Such institutions promote coordination of innovation by providing a forum for collective decision-making among firms, facilitating the introduction of standards (Rysman and Simcoe (2008). If any patent is incorporated into the standards, the patent owner can gain significant power to control the diffusion of such standards and even deter market entry (Shapiro 2001, Rysman and Simcoe 2008). To prevent this blocking effect, most SSOs require patent holders contributing to the standard to license their patents on "Fair, Reasonable, and Non-Discriminatory (FRAND)" terms. Firms can even choose to license their patents with FRAND and royalty-free terms. We control for the incidence of SSO patents that are licensed royalty-free because we expect that such patents might also have some effect on OSS entry. Therefore, we collect all patents disclosed under royalty-free licenses by the major eight SSOs (e.g., IEEE, ITU) from 1971 to 2008 (Rysman and Simcoe 2008)[30]. We compute the claims-weighted patent count of the SSO patents that are distributed under royalty-free licenses and are related to software segment j cumulated up to year t.

6. Empirical Strategies and Results

We motivate our empirical analyses by first investigating the value of patents in the OSS patent pool. If the patents in the pool are not valuable, then they will have little influence on start-up behavior. Next, we establish a relationship between the changes in OSS entry by start-up firms and the changes in the size of the OSS patent pool. We initially assume that any changes in the number of pool patents are uncorrelated with unobservables influencing new OSS entry; we then relax this assumption through instrumental variables estimation. Finally we show that the marginal impact of the OSS patent pool is greater in segments where the cumulativeness of innovation is high or the concentration of patent ownership is high.

6.1 Are the patents in the OSS patent pool less valuable on average than comparable patents?

In our first set of analyses, we examine the quality of patents in the pool relative to a comparison group. Following the matching method employed by Jaffe et al. (1993) and used by many others, we construct a comparison group of patents by choosing the non-pool patents that belong to the same three-digit patent US class as each pool patent and were granted either in the 2 years before the grant year or in the 2 years after the grant year of each pool patent. Table 2 presents how patents in the pools compare to comparable patents with respect to forward citations, backward citations, and number of claims. As shown in the first row of Table 2, pool patents' forward citations are significantly higher than those of control patents. However, non-pool control patents have a significantly higher number of backward citations and claims.

---

[30] We are grateful to the generous offer of the SSO patent data set by Tim Simcoe and Christian Catalini. These data are available for download under a creative commons license at www.ssopatents.org.

This comparison may suggest that while pool patents may cover a narrower technology scope than similar non-pool patents, the pool patents are indeed valuable in the sense that they are less derivative than other comparable patents and are cited more frequently. [31]

6.2. Does the OSS patent pool encourage OSS entry by start-ups?

Proposition 1 shows how changes in the size of an OSS patent pool influence the threshold for entry. The testable implication in our data is how changes in the size of the pool influence the number of entrants. 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 2011). There are several identification challenges in interpreting a relationship between OSS pool patents and OSS entry as a causal one. First, the entry rate and the stock of pool patents may be co-determined by some unobserved segment-specific factor such as variance in demand across different market segments or the stage of the industry life cycle. To mitigate this concern, 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 patent thicket variables. That is, we model new product entry using 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 with parameter λ_{jt} taking the form $\lambda_{jt} = \exp(X_{jt}'\beta)$. Also suppose α_j is a segment-specific and time-constant variable that incorporates unobserved heterogeneity across segments. Thus, $E(Y_{jt} \mid X_{jt}, \alpha_j) = \lambda_{jt} = \alpha_j \exp(X_{jt}'\beta)$, and we assume

$$X_{jt}'\beta = \beta_1 OSS\ patent\ pool_{jt} + \gamma_1 Sales_{jt} + \gamma_2 SegmentPatents_{jt-1} + \gamma_3 PatentThicket_{jt-1}$$
$$+ \gamma_4 OtherFreePatents_{jt} + \tau_t \tag{10}$$

The vector $SegmentPatents_{jt-1}$ includes the *Total patents*$_{jt-1}$, *Patent quality*$_{jt-1}$, and *Potential licensors*$_{jt-1}$; the vector $PatentThicket_{jt-1}$ includes our two patent thicket variables *cumulativeness*$_{jt-1}$ and *concentration*$_{jt-1}$. The two vectors are lagged by one year to allow for any lagged effects on OSS entry. The vector *OtherFreePatents*$_{jt}$ represents the patents from OIN and SSOs—*OIN patents*$_{jt}$ and *SSO patents*$_{jt}$. τ_t includes 10 year dummies to control for time-varying factors that may influence OSS entry. The model is then estimated using maximum likelihood with robust standard errors clustered at the segment level. We are interested in the estimate for β_1 which, if positive, supports proposition 1. To test the robustness of our results, we use different specifications by adding the above controls incrementally.

Table 3 presents the estimation results for specification (10). Column (1) shows the simplest specification with only segment sales as a control as well as with segment and year fixed effects. We

---

[31] We acknowledge that just like the patents disclosed in SSOs (Rysman and Simcoe 2008), one other potential reason that pool patents' forward citations are larger than those of comparable patents is the effect of the patent pool on improving the awareness of these patents, which makes either OSS or non-OSS participants more likely to cite these patents. Thus, the forward citations could be more a measure of follow-up innovations building upon the pool patents rather than a measure of patent quality.

include sales in all specifications because both demand and market competition within a segment are important determinants of start-up entry.[32] The coefficient in column (1) suggests that a 10% increase in the OSS patent pool's patent claims related to a software segment is associated with a 1.4% increase in OSS entry in that segment. Results are robust when we add controls such as the segment's patent size and quality (the vector $SegmentPatents_{jt-1}$), the segment's patent thicket density (the vector $PatentThicket_{jt-1}$), and the size of the patents included in other patent pool-like institutions (the vector $OtherFreePatents_{jt}$). We note that while increases in the size of the OSS patent pool are associated with OSS entry, increases in SSO patents are not. We speculate that this may reflect differences in the licensing requirements for patent pool and SSO patents: in particular, while users of the patent pool pledge not to sue the pool's beneficiaries, licensees of SSO patents have no such requirements. Licensees of SSO patents may see the value of complementary IPRs increase in value, which may increase their incentives to defend their technologies more aggressively. Thus, increases in SSO patents may not reduce the costs of OSS entry.

One potential concern with the above specification is that there may exist unobserved changes specific to the software segment that are correlated both with contributions to the pool and OSS entry. One such possibility is if product market growth is inadequately controlled for by our *Sales* variable. To address this concern, we use a moment-based count data model with instrumental variables and solve the moment conditions through Generalized Method of Moments (GMM) estimation. That is, since the conditional mean $E(Y_{jt} \mid X_{jt}, \alpha_j)$ is equal to $\alpha_j exp(X_{jt}'\beta)$, it implicitly defines the following regression model (Windmeijer and Santos Silva 1997):

$$Y_{jt} = \alpha_j exp(X_{jt}'\beta) + u_{jt} = \alpha_j exp(\beta_1 OSS\ patent\ pool_{jt} + \gamma_1 Sales_{jt} + \gamma_2 SegmentPatents_{jt-1}$$

$$+ \gamma_3 PatentThicket_{jt-1} + \gamma_4 OtherFreePatents_{jt} + \tau_t) + u_{jt} = \mu_{jt}\alpha_j + u_{jt} \qquad (11)$$

where $\mu_{jt} = exp(\beta_1 OSS\ patent\ pool_{jt} + \gamma_1 Sales_{jt} + \gamma_2 SegmentPatents_{jt-1} + \gamma_3 PatentThicket_{jt-1} + \gamma_4 OtherFreePatents_{jt} + \tau_t)$ and u_{jt} is the error term. *OSS patent pool*$_{jt}$ is treated as a potentially endogenous variable. Suppose $X_{jt}\gamma = \gamma_1 Sales_{jt} + \gamma_2 SegmentPatents_{jt-1} + \gamma_3 PatentThicket_{jt-1} + \gamma_4 OtherFreePatents_{jt} + \tau_t$, then X_{jt} is assumed to be exogenous. Following Windmeijer (2000) and Kim and Marschke (2005), we use Wooldridge's quasi-differencing transformation (Wooldridge 1997) to remove the segment-specific fixed effect, and obtain the following moment condition:

$$E\left(Z_{jt}\left(\frac{Y_{jt}}{\mu_{jt}} - \frac{Y_{jt-1}}{\mu_{jt-1}}\right)\right) = 0 \qquad (12)$$

where Z_{jt} includes the set of exogenous variable X_{jt} and a set of instruments for *OSS patent pool*$_{jt}$ as detailed below. As noted by Wooldridge (1997), one drawback for this moment condition is that the estimates of the associated coefficients tend to go infinity if the explanatory variables contain only

---

[32] We have experimented with other controls for market demand such as the number of incumbents. Regressions using these other controls yield qualitatively similar results for the main parameters of interest.

nonnegative values, as is the case for our data. One solution to this problem proposed by Windmeijer (2000) is to transform Z_{jt} as deviations from the overall sample mean; therefore, we transform all Z_{jt} to $Z_{jt} - \bar{z}$, where $\bar{z} = \frac{1}{NT}\sum_{j=1}^{N}\sum_{t=1}^{T} z_{jt}$.

Our first instrumental variable (IV) is the pre-sampling stock of patents at the end of 1997 held by the major contributor to the Patent Commons—IBM—across the 33 software segments. This instrument is designed to capture IBM's propensity to pledge patents across different segments. We expect that incumbents tend to contribute patents to segments where they have accumulated stocks of patents for blocking rivals and facilitating negotiations. A shortcoming with this variable is that its variance is cross-sectional, which would result in its being eliminated by our quasi-differencing method. Thus, we further interact the pre-sampling stock of patents with two time dummies associated with the formation of the OSS patent pool. The first time dummy (denoted as *year2003_2004*) is turned on for year 2003 and year 2004. The motivation to use this time dummy is driven by the observation that on March 7, 2003 IBM was sued by the SCO Group, which asserted that the Linux system embedded by IBM infringed on SCO's UNIX System V source code. This was the first major IPR enforcement lawsuit targeting OSS that attracted significant publicity and as such, it is expected to influence IBM's patent contribution decision. The second time dummy (denoted as *afteryear2005*) is set to be equal to 1 after 2005. This time dummy is designed to reflect the concentration of patent-pledging events for the OSS community during 2005. As described in the Appendix B and also shown by Alexy and Reitzig (2011), following the publication of a report by Open Source Risk Management (OSRM) that Linux had been infringing 283 patents, a series of non-assertion announcements by software industry incumbents came out beginning in 2005. Further, as suggested by Alexy and Reitzig (2011), these incumbents had strong incentives to coordinate with each other to avoid the hold-up problems for the OSS community, resulting in a cluster of non-assertion announcements. Thus, we believe that interacting the pre-sampling stock of IBM patents with the timing of these two events will capture any variance in the propensity of IBM to contribute patents to this pool.

Our second instrumental variable is the cumulated number of IBM patents that were opposed at the European Patent Office (EPO) up to year *t* and related to software segment *j* (denoted as *IBM patents opposed at EPO*). The logic to this instrument is that firms will contribute patents that have the potential to block OSS innovation. Therefore, measures that are correlated with the propensity for IBM to create blocking patents related to a software product market segment over time are, therefore, potential instruments. We argue that one proxy for this propensity is IBM's patents that are opposed at the European Patent Office (EPO). Different from US legal procedures, the EPO allows patents granted at the EPO to be opposed up to nine months after the grant date and at a much lower cost than that of patents opposed through formal legal procedures. Using this instrument has two advantages. First, since the opposition is filed at the EPO, it should not be correlated with new OSS product entries in the US market.

Second, since the opposition needs to be filed within 9 months from the grant of a patent, the truncation issue associated with the lag between the grant date of a patent and the timing of its impact on innovation is far less serious than it is with other similar procedures such as litigation events.[33]

Table 4 presents the GMM estimation results where we instrument for *OSS patent pool*$_{jt}$ using the above instruments. To improve identification, we also add the square of the instruments described above (Gallant 1987). We also test for the validity of all instruments used. As before, we use different specifications by adding the controls incrementally. That is, column (1) in Table 4 provides the estimates using only sales as controls; column (2) further adds the vector *SegmentPatent*; column (3) further adds the vector *PatentThicket*[34]. As we can see from column (1) to column (3), the estimated direct impact of the OSS patent pool is consistently and significantly positive across all specifications, though inclusion of *SegmentPatent* and *PatentThicket* as controls reduces the magnitude of the effect. Note that since the use of the conditional fixed effects model excludes market segments without OSS entry over the sample period, we also present comparable GMM estimates obtained by dropping these segments. The results are presented in column (4) to column (6) of Table 4, and are consistent and similar to the estimates based on the full sample.

We probe the validity of our instruments by first examining the power of the instruments, by using the F-test from an auxiliary first stage ordinary least square (OLS) regression of the endogenous variable against the same set of IVs and exogenous variables. As shown in Table A-4 in Appendix D, the F-statistics on the coefficients of our instruments range from 35.17 to 83.63, depending upon the specification, and are all statistically significant, which suggests that the instruments have some power in explaining the endogenous variable.[35] We also perform the Hansen J statistic to test the over-identification restrictions. The results of these tests are in the last row of Table 4. All tests fail to reject the null that the instruments used are uncorrelated with the error term across all specifications. We also test the validity of a subset of the instruments. In particular, we assume that the stock of IBM patents opposed at EPO is exogenous, and test the validity of the interactions between the pre-sampling stock of IBM patents with the year dummies.[36] As shown in Table A-4, based on the C-test statistic we fail to reject the null that the

---

[33] More specifically, as has been noted by Hall et al. (2003), because oppositions must be filed within 9 months of a patent grant, the average lag between applying for a patent and the filing of opposition request is relatively tight; in contrast, other legal procedures such as litigation and patent re-examinations can be initiated at any time during the lifetime of a patent, which results in greater variance in the distribution of these procedural lags.

[34] We were not able to add the vector *OtherFreePatents*, as it leads to non-convergent results.

[35] As an additional robustness check, another set of IVs has been constructed: we interact the pre-sampling stock of IBM patents with a *year 2003* dummy (which is set to be 1 for year 2003 and 0 for other years) and an *afteryear2004* dummy (which is set to be 1 after year 2004 and 0 for other years); we use these two plus their squares and IBM opposed patents at EPO plus its square as IVs for GMM estimation using the same model. The results are similar and presented in Table A-6.

[36] We implement a C-test to evaluate the exogeneity of this subset of IVs, with the null hypothesis that they are valid instruments. That is, we treat the full model with the three IVs and their squares as the restricted and fully efficient regression; we use the model with the IBM patents opposed at EPO and its square as IVs as the unrestricted, inefficient but consistent regression.

IVs—the *pre-sampling stock of IBM patents* times *year2003_2004* and its square and the *pre-sampling stock of IBM patents* times *afteryear2005* and its square—are valid across all specifications.[37]

Finally, to further boost confidence in our results, we also implement a falsification exercise. The intuition is that, given our theoretical predictions, we should not observe a positive effect of the OSS pool on start-up entry based on proprietary software product entry. Focusing on PROMT articles with PROMT product codes, we identified entry with proprietary product introductions by start-ups into 29 software segments from year 2002 to year 2009.[38] As shown in Table A-5 in Appendix D, our results suggest that there is no significantly positive effect of the pool on entry using this set of products.

6.3. How does the marginal effect of the OSS patent pool vary with the characteristics of the patent thicket?

In this section, we investigate how the impact of the OSS patent pool varies with the cumulativeness of innovation and the concentration of patent ownership in a market segment. We begin by examining whether the impact of the OSS patent pool is higher when the cumulativeness of innovation in a segment is high. The specification for $X_{jt}'\beta$ becomes

$$X_{jt}'\beta = \beta_1 OSS\ patent\ pool_{jt} + \beta_2 OSS\ patent\ pool_{jt}*cumulativeness_{jt-1}$$

$$+ \gamma_1 Sales_{jt} + \gamma_2 SegmentPatents_{jt-1} + \gamma_3 PatentThicket_{jt-1} + \gamma_4 OtherFreePatents_{jt} + \tau_t. \quad (13)$$

To test the direct impact of the OSS patent pool (proposition 1), we present the marginal effect of the OSS patent pool when other variables are at their mean level. Our propositions 2 and 3 state how changes in patent pools, cumulativeness, and concentration influence the threshold for entry. The testable implications in our data are how changes in each of these variables influence the number of entrants. Thus, if the marginal effect of patent pools on entry is greater when cumulativeness is high, this is supportive of proposition 2. Similarly, if the marginal effect of patent pools on entry is greater when concentration is high, this is supportive of proposition 3. To capture the interaction effects as suggested by proposition 2, we compute the marginal effect of the pool when cumulativeness is at high and low levels, and test whether the marginal effect of the OSS patent pool is significantly different at these two levels. As before, we employ different specifications by adding the sets of four controls incrementally and use an alternative measure of cumulativeness to probe the robustness of the results.

As shown in columns (1)-(3) in Table 5, a 10% increase in the pool's patent claims is associated with a 2.2%-3% increase in OSS entry, with the effect computed at the average level of cumulativeness of

---

[37] Alternatively, we also performed a test on the validity of the subset of instruments represented by the stock of IBM patents opposed at the EPO, and this time assume that the interactions between the pre-sampling stock of IBM patents and the year dummies are exogenous. Based on the C-test statistic, we fail to reject the null that the stock of IBM patents opposed at the EPO is valid across all specifications.

[38] Our use of the aggregated 29 segments (rather than the baseline 33 segments) from 2002 to 2009 reflects our data constraints. More details are provided in Table A-5 in Appendix D.

innovation. Further, while the marginal effect of the OSS patent pool is insignificant when evaluated at the 10th percentile of the cumulativeness, the effects are statistically and economically significant when evaluated at the 90th percentile. Specifically, a 10% increase in the pool's patent claims is associated with a 3.8%-5.6% increase in OSS entry when cumulativeness of innovation is at its 90th percentile. A test for the difference of the two marginal effects (at the 10th and 90th percentiles) is statistically significant. Meanwhile, all the estimates are stable across all specifications and across different measures of cumulativeness of innovation.[39] These results suggest that the impact of the OSS patent pool is greater when the cumulativeness of innovation is high, providing evidence in support of proposition 2.

Our next step is to explore how the impact of the OSS patent pool is influenced by variation in the concentration of patent ownership. The specification can be written as

$$X_{jt}'\beta = \beta_1 OSS\,patent\,pool_{jt} + \beta_3 OSS\,patent\,pool_{jt} *\,concentration_{jt-1}$$
$$+ \gamma_1 Sales_{jt} + \gamma_2 SegmentPatents_{jt-1} + \gamma_3 PatentThicket_{jt-1} + \gamma_4 OtherFreePatents_{jt} + \tau_t. \qquad (14)$$

The empirical results for this specification are shown in columns (4)-(6) in Table 5. While the marginal effect of the OSS patent pool is insignificant when concentration is at its 10th percentile or mean value, a 10% increase in the OSS patent pool's patent claims is associated with a 1.3%-1.7% increase in OSS entry when concentration is at its 90th percentile. Meanwhile, excluding the simplest specification with only sales as control, the test for the difference of marginal effects of the OSS patent pool between concentration evaluated at the 10th percentile and the 90th percentile is statistically significant at the 10% level. Thus, our results provide evidence in support of proposition 3.

To present a more complete picture of how the impact of the OSS patent pool varies with cumulativeness of innovation and concentration of patent ownership, we present results including the two sets of interactions together. These estimates are presented in columns (7)-(9) in Table 5. A 10% increase in the pool's patent claims is associated with a 4.5%-4.8% increase in OSS entry when the cumulativeness of innovation is at its 90th percentile; the marginal effect of the pool is significantly different at the 1% level when evaluated at high and low levels of cumulativeness. While in this specification the interaction between the OSS patent pool and concentration of patent ownership becomes insignificant, the impact of the OSS patent pool is still statistically significant and positive when evaluated at the 90th percentile of concentration and the sign of the interaction's coefficient is positive across specifications. However, there is no statistically significant difference between the marginal effects evaluated at the 10th and 90th percentiles of concentration. Thus, while the qualitative nature of our results is similar when including

---

[39] The regression results based on the robust measure of cumulativeness are shown in Table A-7 in Appendix D.

cumulativeness and concentration together, the statistical significance of the concentration result is weaker. This is likely caused by the multicollinearity between the two interactions.[40]

To evaluate how potential omitted variables may influence our results, we again examine the robustness of our results to the use of instrumental variables. We interact each of the original instruments (for the size of the OSS patent pool) with cumulativeness and with patent ownership concentration to form the instruments for the two interaction variables. The estimated results are shown in Table 6. Columns (1) through (3) provide estimates using the full sample. The estimated coefficients for the two interaction variables remain significantly positive across all specifications. For robustness, the estimates from columns (4) through (6) are based on a sample that excludes market segments without OSS entry during the study period; the coefficients for the two interactions are still significantly positive and stable across all specifications. While the combined empirical evidence from Tables 5 and 6 largely confirms proposition 3, it seems to suggest that the interaction between the OSS patent pool and concentration of patent ownership is not as strong as its interaction with cumulativeness.

We also explore the impact of the OSS patent pool when both the cumulativeness of innovation and concentration of patent ownership in a segment are high.[41] For segments with both cumulativeness of innovation and concentration at high levels (at the 90th percentile), a 10% increase in the pool's patent claims is associated with a 7.5%-8.8% increase of new OSS entry; this marginal effect is significantly greater than that when cumulativeness of innovation or concentration (or both) are at lower levels.

7. Conclusions

Our empirical evidence demonstrates that increases in the size of an OSS patent pool related to a software segment are associated with increases in OSS entry by start-up software firms in that segment.[42] Furthermore, the impact of the OSS patent pool is magnified when two features of patent thickets are present in the segment: cumulativeness of innovation and concentration of patent ownership. We observe a particularly strong relationship between the size of the patent pool and OSS start-ups' entry in segments with high cumulativeness of innovation; the marginal effect of the pool is also greater when concentration is high, although this result is not robust across all specifications.

In conclusion, our results suggest that OSS patent pools may facilitate markets for technology by strengthening a startup's negotiating position against incumbents with potentially blocking patents. Indeed, by reducing entry costs and the associated incentives for startups to operate in the open source

---

[40] In the pooled sample, the simple correlation coefficient between the two interaction terms is 0.66.

[41] The detailed empirical results are shown in Table A-8 in Appendix D.

[42] A natural and interesting extension to this study is to look at how the size of the OSS patent pool influences the survival rate of OSS firms. However, of the firms introducing new OSS products in our sample, only one firm exited before the end of the sampling period. As a result, there is insufficient variance in our data to measure the impact of the OSS patent pool on the survival of start-up firms who produce OSS.

world, OSS patent pools appear to stimulate the open source innovation activities of entrepreneurial firms in industries characterized by dense patent thickets and concentrated property rights.

Our study analyzes the impact of patent pools on the behavior of those firms whose entry decisions are most likely to be affected by the change in licensing and negotiation costs: start-up firms considering entry as an OSS competitor. The introduction of OSS patent pools may have secondary implications for two groups of firms that we do not study: large firms and those who sell software under a traditional proprietary license. Understanding the implications of OSS patent pools on these other groups will have important implications for the rate and direction of inventive activity in software, and quantifying these implications is an important question for future research.

This study deepens our understanding of the role of patent pools. While prior work has shown that the introduction of traditional patent pools can lead to a decline in innovative activity among both licensors and licensees (Lampe and Moser 2010, Joshi and Nerkar 2011), we find that the introduction of OSS patent pools stimulates the innovative activity of a key group, start-ups. We speculate that this difference may reflect the requirements of use for the OSS patent pool; namely OSS pool patents are offered royalty-free and beneficiaries are required not to sue firms producing OSS. As noted above, the absence of this latter requirement may be one reason for our inability to find a measurable impact of standard setting organizations' patents on OSS entry. However, more research is needed to understand how licensing requirements across the two types of institutions influence innovation outcomes.

References

Alexy, O., Reitzig, M. G. 2011. Private-Collective Innovation, Competition, and Firms' Counterintuitive Appropriation Strategies. Available at SSRN: http://ssrn.com/abstract=1430328

Asay, Matt. 2010. What Apple's and Microsoft's patent threats mean for start-ups. CNET. Available at http://news.cnet.com/8301-13505_3-10466670-16.html

Bessen, J. E. and R. M. Hunt. 2007. An Empirical Look at Software Patents. *Journal of Economics & Management Strategy* 16(1): 157-89.

Bonaccorsi, A., S. Giannangeli, C. Rossi. 2006. Entry Strategies Under Competing Standards: Hybrid Business Models in the Open Source Software Industry. *Management Science* 52(7): 1085-1098.

Broersma, M. 2007. IDC: Open-Source Market to Be Worth $5.8B by 2011. CIO. Available at http://www.cio.com/article/116201/IDC_Open_Source_Market_to_Be_Worth_5.8B_by_2011

Pressual 1993, Vol. 8, pp. 15-86, National Bureau of Economic Research, Cambridge, MA.

Caballero, R.J., A.B. Jaffe. 1993. How High are the Giants' Shoulders: An Empirical Assessment of Knowledge Spillovers and Creative Destruction in a Model of Economic Growth. In O. Blanchard and S. Fischer, eds., National Bureau of Economic Research Macroeconomics Annual, Vol. 8, pp. 15-86, MIT Press, Cambridge, MA.

Clarkson, G. 2005. Patent Informatics for Patent Thicket Detection: A Network Analytic Approach for Measuring the Density of Patent Space. Working Paper. Available at: http://w4.stern.nyu.edu/emplibrary/ACFltbnmV.pdf

Cockburn, I. M., M. MacGarvie. 2006. Entry, Exit and Patenting in the Software Industry. NBER Working paper 12563, National Bureau of Economic Research, Cambridge, MA.

Cockburn, I. M., M. 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., M. MacGarvie. 2011. Entry and Patenting in the Software Industry. *Management Science* 57(5): 915-933.

Dahlander, L. 2007. Penguin in a new suit: a tale of how de novo entrants emerged to harness free and open source software communities. *Industrial and Corporate Change* 16 (5): 913-943.

Dillon, M. 2008. Firestar. Web log comment. Available at http://blogs.oracle.com/dillon/entry/firestar (Accessed December 12, 2011).

Fosfuri, A., M. S. Giarratana, A. Luzzi. 2008. The Penguin Has Entered the Building: The Commercialization of Open Source Software Products. *Organization Science* 19(2): 292-305.

Galasso, A., M. Schankerman. 2010. Patent Thickets, Courts, and the Market for Innovation. *RAND Journal of Economics* 41(3): 472-503.

Gallant, A.R. 1987. *Nonlinear Statistical Models*. John Wiley & Sons, Inc.

Gambardella, A., B. H. Hall. 2006. Proprietary versus public domain licensing of software and research products. *Research Policy* 35(6): 875-892.

Graham, S. J. H., D. 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.

Griliches, Z. 1990. Patent statistics as economic indicators: a survey. *Journal of Economic Literature* 27: 1661–1707.

Hall, B.H., Helmers, C. 2011. Innovation and Diffusion of Clean/Green Technology: Can Patent Commons Help? NBER Working Paper 16920, National Bureau of Economic Research, Cambridge, MA.

Hall, B.H., S. Graham, D. Harhoff. 2003. Prospects for Improving Us Patent Quality Via Post-Grant Opposition. NBER Working Paper 9731, National Bureau of Economic Research, Cambridge, MA.

Hall B.H., A. Jaffe, M. Trajtenberg. 2005. Market value and patent citations. *RAND Journal of Economics* 36 (1): 16 – 38.

Hall, B. H., A. Jaffe, M. Trajtenberg. 2001. The NBER Patent Citation Data File: Lessons, Insights and Methodological Tools. NBER Working Paper 8498, National Bureau of Economic Research, Cambridge, MA.

Jaffe, A.B, M.S. Fogarty, B.A. Banks. 1998. Evidence from Patents and Patent Citations on the Impact of NASA and Other Federal Labs on Commercial Innovation. *Journal of Industrial Economics* 46(2): 183-205.

Jaffe A., M. Trajtenberg, R. Henderson. 1993. Geographic localization of knowledge spillovers as evidenced by patent citations. *The Quarterly Journal of Economics* 108(3): 577-598.

Joshi, A.M., A. Nerkar. 2011. When Do Strategic Alliances Inhibit Innovation by Firms? Evidence from Patent Pools in the Global Optical Disc Industry. *Strategic Management Journal* 32(11): 1139-1160

Kim, J., G. Marschke. 2005. Labor Mobility of Scientists, Technological Diffusion, and the Firm's Patenting Decision. *RAND Journal of Economics* 36(2): 298-317.

Lampe, R., P. Moser. 2010. Do Patent Pools Encourage Innovation? Evidence from the Nineteenth-Century Sewing Machine Industry. *The Journal of Economic History* 70(4): 898-920.

Layne-Farrar, A., J. Lerner. 2010. To Join or Not to Join: Examining Patent Pool Participation and Rent Sharing Rules *International Journal of Industrial Organization* 29(2) 294-303.

Lerner, J., M. Schankerman. 2010. *The Comingled Code: Open Source and Economic Development*. Cambridge: MIT Press.

Lerner, J., M. Strojwas, J. Tirole. 2007. The Design of Patent Pools: The Determinants of Licensing Rules. *The RAND Journal of Economics* 38(3): 610-25.

Lerner, J., J. Tirole. 2004. Efficient patent pools. *American Economic Review* 94(3): 691-711.

Lerner, J., J. Tirole. 2005a. The Economics of Technology Sharing: Open Source and Beyond. *Journal of Economic Perspectives* 19 (Spring): 99-120.

Lerner, J., J. Tirole. 2005b. The Scope of Open Source Licensing. *Journal of Law, Economics, and Organization* 21(April): 20-56.

Lerner, J., F. Zhu. 2007. What Is the Impact of Software Patent Shifts? Evidence from Lotus V. Borland. *International Journal of Industrial Organization* 25(3): 511-29.

Lévêque, F., Ménière, Y. 2007. Copyright versus Patents: The Open Source Software Legal Battle. *Review of Economic Research on Copyright issues* 4(1): 27-46.

Llobet, G. 2003. Patent Litigation When Innovation is Cumulative. *International Journal of Industrial Organization* 21: 1135-1157.

Maurer, S. M., S. Scotchmer. 2006. Open Source Software: The New Intellectual Property Paradigm. Hendershott, T eds. *Handbook of Economics and Information Systems* 285-319, Elsevier.

Merges, R. P., R. R. Nelson. 1990. On the Complex Economics of Patent Scope. Columbia Law Review. 90(4) 839-916.

Noel, M.D., M.A. Schankerman. 2006. Strategic Patenting and Software Innovation: Theory and evidence from a panel of U.S. firms. CEPR Working Paper 5701, Center for Economic Policy Research, Washington, DC.

O'Donoghue, T., S. Scotchmer, and J.-F. Thisse. 1998. Patent Breadth, Patent Life, and the Pace of Technological Progress. *Journal of Economics and Management Strategy* 7(1): 1-32.

Patent Commons project. 2005. About Commitments. Available at http://www.patent-commons.org/resources/about_commitments.php

Paul, R. 2008a. Sun eclipses Red Hat Firestar pact as patent invalidated. Web log comment. Available at http://arstechnica.com/open-source/news/2008/07/sun-eclipses-red-hat-firestar-pact-as-patent-invalidated.ars (Accessed December 12, 2011).

Paul, R. 2008b. Barracuda bites back at Trend Micro in ClamAV patent lawsuit. Web log comment. Available at http://arstechnica.com/open-source/news/2008/07/barracuda-bites-back-at-trend-micro-in-clamav-patent-lawsuit.ars (Accessed December 12, 2011).

Rysman, M., T. Simcoe. 2008. Patents and the Performance of Voluntary Standard Setting Organizations. *Management Science* 54(11): 1920-1934.

Schankerman, M. and S. Scotchmer. 2001. Damages and injunctions in protecting intellectual property. *RAND Journal of Economics* 32(1): 199-220.

Scotchmer. 2005. *Innovation and Incentives*. MIT Press, Cambridge, MA.

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

Serafino, D. 2007. Survey of Patent Pools Demonstrates Variety of Purposes and Management Structures. KEI Research Note 6.

Shapiro, C. 2001. Navigating the Patent Thicket: Cross Licenses, Patent Pools, and Standard-Setting. A. Jaffe, J. Lerner, S. Stern, eds. *Innovation Policy and the Economy, Vol. 1*. MIT Press, Cambridge, MA.

Silverman, B.S. 1999. Technological Resources and the Direction of Corporate Diversification: Toward an Integration of the Resource-Based View and Transaction Cost Economics. *Management Science* 45(8): 1109-24.

von Hippel, E., G. von Krogh. 2003. Open Source Software and the 'Private-Collective' Innovation Model: Issues for Organization Science. *Organization Science* 14(2): 209-223.

Windmeijer, F. 2000. Moment Conditions for Fixed Effects Count Data Models with Endogenous Regressors. *Economics Letters* 68(1): 21-24.

Windmeijer, F., J. Santos Silva. 1997. Endogeneity in Count Data Models: An Application to Demand for Health Care. *Journal of Applied Econometrics* 12(3): 281-294.

Wooldridge, J.M. 1997. Multiplicative Panel Data Models without the Strict Exogeneity Assumption. *Econometric Theory* 13: 667-678.

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.

<h2 style="text-align:center">Table 1: Summary Statistics</h2>

Variable name	Measure (Segment-year)	Obs.	Mean	Std. Dev.	Min	Max
Dependent variable						
OSS entry	The number of new OSS entrants into segment j in year t	363	.667	1.344	0	11
Independent variables and controls						
OSS patent pool	Log of Patent Commons' claims-weighted patent count related to segment j cumulated up to year t	363	2.413	2.936	0	7.911
Cumulativeness	Log of cumulativeness of innovation in segment j up to year t-1	363	.808	.613	.095	3.454
Cumulativeness_R	Log of cumulativeness of innovation in segment j up to year t-1 using the robustness measure proposed by Clarkson (2005)	363	.473	.490	.036	2.933
Concentration	Four-assignee citation concentration ratio in segment j up to year t-1	363	0.227	0.075	0.076	0.458
Potential licensors	Log of total number of assignees (divided by 100) that are cited by patents in segment j up to year t-1	363	5.559	1.301	1.072	8.407
OIN patents	Log of Open Invention Network's claims-weighted patent count in segment j cumulated up to year t	363	1.125	2.053	0	6.690
SSO patents	Log of standard-setting organizations' claims-weighted patent count in segment j cumulated up to year t	363	1.628	2.118	0	5.908
Total patents	Log of total claims-weighted patent count related to segment j cumulated up to year t-1	363	10.817	1.232	6.870	13.486
Patent quality	Log of average quality of patents related to segment j cumulated up to year t-1	363	2.832	.402	1.839	4.051
Sales	Log of total volume of sales by segment j in year t (in Million)	363	7.223	1.523	2.660	9.145
Instrument variables						
IBM patents * year2003_2004	Log of IBM's pre-sampling stock of claims-weighted patent count * year2003_2004 dummy, where year2003_2004 dummy turns on for year t = 2003, 2004	363	1.209	2.708	0	9.838
IBM patents * after_year2005	Log of IBM's pre-sampling stock of claims-weighted patent count * after_year2005 dummy, where after_year2005 dummy turns on for year t = 2005, 2006, …, 2009	363	3.023	3.584	0	9.838
IBM patents opposed at EPO	Log of IBM's patents granted by the European Patent Office (EPO) and opposed at EPO related to segment j cumulated up to year t	363	1.596	.829	0	3.367
IBM patents * year 2003	Log of IBM's pre-sampling stock of claims-weighted patent count * year2003 dummy, where year2003 dummy turns on for year t = 2003	363	.605	2.009	0	9.838
IBM patents * after_year2004	Log of IBM's pre-sampling stock of claims-weighted patent count * after_year2004 dummy, where after_year2004 dummy turns on for year t = 2004, 2005, …, 2009	363	3.627	3.635	0	9.838

Table 2: Pool Patents Compared to Non-pool Patents

		Pool Patents	Non-pool Control Patents	T-test
	Obs	2117	250618	
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 truncation based on the methods by Hall et al. (2001); 2) Following the matching method employed by Jaffe et al. (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 three-digit class as each of the pool patents and were granted either in the 2 years before the grant year or in the 2 years after the grant year of each pool patent; 3) ***: significant at 1%.

Table 3. Conditional Fixed-effect Poisson Regression: Direct Impact of an OSS Patent Pool

Dependent variable: OSS entry			
	(1)	(2)	(3)
OSS patent pool	.141* (.081)	.160** (.087)	.154* (.091)
Sales	-.003 (.611)	-.045 (.586)	-.045 (.536)
Potential licensors		.715 (1.697)	1.217 (1.654)
Total patents		-1.770 (2.427)	-1.895 (2.527)
Patent quality		-1.140 (2.391)	-1.242 (2.391)
OIN patents			-.057 (.076)
SSO patents			-.015 (.074)
Cumulativeness			1.152 (1.134)
Concentration			-9.348 (6.818)
Observations	286	286	286
Log pseudolikelihood	-229.734	-229.201	-227.777

Notes: 1) Robust standard errors, clustered by market segment, are in parentheses. 2) * significant at 10%, ** significant at 5%, *** significant at 1%. 3) The number of observations is lower than 363 (i.e., the number of observations listed in Table 1) because of the use of conditional fixed effects Poisson models, which drops market segments without OSS entry over the entire sample period. 4) All regressions include market and year fixed effects.

Table 4. GMM Estimates of Count Data Regression with IVs: Direct Impact of an OSS Patent Pool

Dependent variable: OSS entry	Full sample			Sample dropping the segments without OSS entry		
	(1)	(2)	(3)	(4)	(5)	(6)
OSS patent pool	.430***	.241*	.249**	.406***	.258**	.212*
	(.139)	(.127)	(.102)	(.153)	(.132)	(.111)
Sales	.100	-.178	-.397*	.132	-.128	-.298
	(.345)	(.153)	(.215)	(.348)	(.149)	(.217)
Potential licensors		2.094	3.260		1.907	2.782
		(2.092)	(2.045)		(2.142)	(2.270)
Total patents		-1.750	-2.773		-1.537	-2.455
		(2.186)	(1.872)		(2.220)	(2.037)
Patent quality		1.212	2.764***		1.464*	2.853***
		(.843)	(.869)		(.757)	(.903)
Cumulativeness			-.404			-.700
			(.924)			(1.119)
Concentration			-8.671			-7.488
			(6.961)			(7.213)
Year dummies	Yes	Yes	Yes	Yes	Yes	Yes
Observations	363	363	363	286	286	286
Over-identification test statistic (p-value)	10.303	15.313	17.442	10.521	15.030	17.097
	(0.24)	(0.17)	(0.10)	(0.23)	(0.18)	(0.11)

Notes: 1) The full set of IVs include IBM patents*year2003_2004, IBM patents*afteryear2005, square of IBM patents*year2003_2004, square of IBM patents*afteryear2005, IBM opposed patents at EPO, and square of IBM opposed patents at EPO. 2) Robust standard errors, clustered by market segment, are in parentheses. 3) * significant at 10%, ** significant at 5%, *** significant at 1%. 4) Year dummies include year2001_2002, year2003, year2004, year2005, year2006, year2007, and year2008_2009. The reason for not using the full set of ten year dummies is that including more year dummies leads to non-convergence of the GMM estimator.

Table 5. Conditional Fixed-effect Poisson Regression: Direct Impact of an OSS Patent Pool and its Interaction with Cumulativeness of Innovation and with Concentration of Patent Ownership

Dependent variable: OSS entry	Interaction with Cumulativeness of Innovation			Interaction with Concentration of Patent Ownership			Interaction with Cumulativeness of Innovation and with Concentration of Patent Ownership		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
OSS patent pool	.079 (.118)	.019 (.103)	.065 (.134)	-.012 (.140)	-.012 (.130)	-.080 (.145)	-.082 (.144)	-.066 (.142)	-.048 (.182)
OSS patent pool * Cumulativeness	.191* (.102)	.265*** (.098)	.314*** (.096)				.297*** (.090)	.289*** (.092)	.290*** (.098)
OSS patent pool * Concentration				.422 (.351)	.541* (.322)	.640* (.381)	.287 (.301)	.310 (.280)	.312 (.386)
Cumulativeness	1.006* (.557)	2.722*** (.758)	3.225*** (1.027)			1.184 (1.019)	3.057*** (.889)	2.934*** (.865)	3.081*** (1.036)
Concentration			-13.901** (6.896)	-9.417 (6.891)	-12.958 (6.628)	-12.632* (7.308)	-14.150*** (5.362)	-14.829*** (5.672)	-15.093** (6.799)
Sales	-.158 (.545)	-.048 (.534)	-.018 (.501)	-.139 (.565)	-.038 (.555)	-.088 (.522)	-.044 (.519)	-.036 (.524)	-.036 (.498)
Potential licensors		0.501 (1.658)	0.965 (1.573)	-.710* (.438)	1.113 (1.645)	1.367 (1.652)	.961 (.643)	1.027 (1.636)	1.038 (1.566)
Total patents		.512 (2.384)	-.317 (2.532)		-3.112 (2.534)	-2.514 (2.610)		-.845 (2.538)	-.700 (2.574)
Patent quality		-.886 (2.333)	-1.176 (2.377)		-1.388 (2.483)	-1.329 (2.418)		-1.183 (2.424)	-1.203 (2.393)
OIN patents			-.059 (.069)			-.067 (.074)			-.062 (.069)
SSO patents			-.020 (.062)			.023 (.079)			-.001 (.075)
Observations	286	286	286	286	286	286	286	286	286
Log pseudolikelihood	-227.205	-225.971	-224.103	-228.567	-227.599	-226.762	-224.319	-224.186	-223.893
Marginal Effects									
OSS patent pool (average)	.221** (.103)	.216** (.094)	.298** (.128)	.081 (.090)	.108 (.087)	.062 (.086)	.202** (.010)	.217** (.101)	.236** (.126)
OSS patent pool (cumulativeness=10%)	.121 (.108)	.077 (.094)	.133 (.129)				.046 (.108)	.066 (.107)	.084 (.124)
OSS patent pool (cumulativeness=90%)	.383*** (.146)	.440*** (.140)	.564*** (.164)				.453*** (.128)	.462*** (.135)	.481*** (.166)
Statistic for the test of the difference between high and low cumulativeness (p-value)	3.49* (0.06)	7.32*** (0.01)	10.71*** (0.00)				10.91*** (0.00)	9.79*** (0.00)	8.82*** (0.00)
OSS patent pool (concentration =10%)				.043 (.107)	.059 (.101)	.003 (.107)	.176 (.110)	.189* (.111)	.207 (.148)
OSS patent pool (concentration =90%)				.129* (.083)	.169** (.081)	.134* (.076)	.234** (.096)	.253*** (.097)	.271** (.109)
Statistic for the test of the difference between high and low concentration (p-value)				1.45 (0.20)	2.83* (0.09)	2.82* (0.09)	0.91 (0.34)	1.23 (0.26)	0.66 (0.40)

Notes: 1) Robust standard errors, clustered by market segment, are in parentheses. 2) * significant at 10%, ** significant at 5%, *** significant at 1%. 3) The number of observations is lower than in Table 1 because of the use of conditional fixed effects Poisson models, which drop market segments in which there is no new product entry over our sample period. 4) The statistic for the test of the difference of marginal effect of an OSS patent pool between the cumulativeness of innovation/the concentration of patent ownership at the 10th percentile and at the 90th percentile is distributed as chi-square, and the p-value is shown in the parentheses. 5) All regressions include market and year fixed effects.

Table 6. GMM Estimates of Count Data Regression with IVs: Direct Impact of an OSS Patent Pool and its Interactions with Cumulativeness of Innovation and with Patent Ownership Concentration

Dependent variable: OSS entry	Full sample			Sample dropping the segments without OSS entry		
	(1)	(2)	(3)	(4)	(5)	(6)
OSS patent pool	-.432***	-.679***	-.478	-.401***	-1.070***	-.865**
	(.130)	(.201)	(.341)	(.127)	(.302)	(.404)
Cumulativeness	1.956***	.587	.697	2.176***	.536	1.340
	(.564)	(1.629)	(2.737)	(.592)	(2.110)	(1.222)
OSS patent pool * Cumulativeness	.567***	.605***	.583**	.588***	.462*	.489*
	(.138)	(.210)	(.265)	(.141)	(.256)	(.283)
Concentration	-15.717***	-20.963***	-18.632	-14.483***	-21.017***	-17.147***
	(4.114)	(5.950)	(6.287)	(4.133)	(5.651)	(5.527)
OSS patent pool * Concentration	2.338***	2.161***	1.626*	1.976***	2.746***	1.973*
	(.417)	(.604)	(.870)	(.376)	(.626)	(1.019)
Potential licensors	1.482***	-1.036	-.368	1.551***	.291	.888
	(.355)	(1.366)	(1.858)	(.371)	(1.225)	(1.550)
Sales	-.280	-.267	-.232	-.245	-.145	-.186
	(.261)	(.245)	(.242)	(.277)	(.247)	(.232)
Total patents		2.623*	1.977		1.285	1.054
		(1.547)	(1.689)		(1.698)	(1.861)
Patent quality		4.661***	4.293***		4.669***	4.007***
		(1.206)	(1.464)		(1.138)	(1.206)
SSO patents			-.134			-.139
			(.105)			(.106)
Year dummies	Yes	Yes	Yes	Yes	Yes	Yes
Observations	363	363	363	286	286	286
Over-identification test statistic (p-value)	19.425 (0.08)	16.201 (0.18)	16.377 (0.18)	19.182 (0.08)	16.622 (0.22)	17.978 (0.16)

Notes: 1) The full set of IVs include: IBM patents*year2003_2004, IBM patents*afteryear2005, square of IBM patents*year2003_2004, square of IBM patents*afteryear2005, IBM opposed patents at EPO, and square of IBM opposed patents at EPO. 2) Robust standard errors, clustered by market segment, are in parentheses. 3) * significant at 10%, ** significant at 5%, *** significant at 1%. 4) Year dummies include year2001_2002, year2003, year2004, year2005, year2006, year2007, and year2008_2009. Including more year dummies would lead to non-convergence of the GMM estimator. 5) We are unable to include OIN patents as a control, as it would also lead to non-convergent estimation results.

APPENDIX A: Proofs in Section 2

1. Proof for Lemma 1 As assumed above, Δ is distributed uniformly and $g(\Delta) = 1$, so $G(\widetilde{\underline{\Delta}}) = \widetilde{\underline{\Delta}}$, $G(\Delta_s(T)) = \Delta_s(T)$. Given $q(\Delta, b) = b(1 - \Delta)$, we have

$$\max_T \pi(v, b, L|\widetilde{\underline{\Delta}}) = \max_T \int_{\Delta_s(T)}^{1}(b(1 - \Delta)v - L)\frac{1}{1-\widetilde{\underline{\Delta}}}d\Delta + T \cdot \frac{\Delta_s(T)-\widetilde{\underline{\Delta}}}{1-\widetilde{\underline{\Delta}}}$$

$$= \max_T \frac{1}{1-\widetilde{\underline{\Delta}}}\int_{\Delta_s(T)}^{1}(bv - bv\Delta - L)\,d\Delta + T \cdot \frac{\Delta_s(T)-\widetilde{\underline{\Delta}}}{1-\widetilde{\underline{\Delta}}}$$

Let's denote $\Delta_s(T)$ as Δ_s and since it is more convenient we compute the first order condition with respect to Δs,

$$\max_{\Delta_s} \pi(v, b, L|\widetilde{\underline{\Delta}}) = \max_{\Delta_s} \frac{1}{1-\widetilde{\underline{\Delta}}}\left(bv - \frac{1}{2}bv - L - bv\Delta_s + \frac{\Delta_s^{2}}{2}bv + \Delta_s L\right) + \frac{\Delta_s-\widetilde{\underline{\Delta}}}{1-\widetilde{\underline{\Delta}}}\cdot(b(1 - \Delta_s)v + L)$$

$$\frac{\partial\pi}{\partial\Delta_s} = \frac{1}{1-\widetilde{\underline{\Delta}}}(-bv + \Delta_s bv + L) + \frac{b(1-\Delta_s)v+L}{1-\widetilde{\underline{\Delta}}} + \frac{\Delta_s-\widetilde{\underline{\Delta}}}{1-\widetilde{\underline{\Delta}}}\cdot(-vb)$$

The first order condition $\frac{\partial\pi}{\partial\Delta_s} = 0$ is equal to $-\Delta_s bv + 2L + vb\widetilde{\underline{\Delta}} = 0$.

That is, the optimal Δ^s for a given correct belief $\widetilde{\underline{\Delta}}$ is:

$$\Delta^{s*} = \frac{1}{bv}(2L + vb\widetilde{\underline{\Delta}}) \tag{15}$$

Since $T = b(1 - \Delta_s)v + L$, the corresponding T for a given correct belief $\widetilde{\underline{\Delta}}$ would be

$$T = bv - (2L + vb\widetilde{\underline{\Delta}}) + L \tag{16}$$

The unique sequential equilibrium in pure strategies can be obtained from the following two conditions, where the first condition (equation (17)) describes the optimal $T(v, b, L|\widetilde{\underline{\Delta}})$ given a $\widetilde{\underline{\Delta}}$ and the second condition (equation (18)) describes how the decision of OSS entry is determined by a given T:

$$T^* = bv - (2L + vb\underline{\Delta}^*) + L \tag{17}$$

$$\underline{\Delta}^* = T^* + c - v \tag{18}$$

Combining these two conditions, we have

$$\underline{\Delta}^* = vb - (2L + vb\underline{\Delta}^*) + L + c - v$$

Therefore, the optimal improvement level $\underline{\Delta}^*$ for entry has the following form:

$$\underline{\Delta}^* = (1 + vb)^{-1}(vb + c - L - v) \tag{19}$$

Further, to satisfy $\underline{\Delta}^* \in [0,1]$, $vb + c - L - v < 1 + vb$, i.e. $L + v + 1 - c > 0$ (20)

2. Proof for Proposition 1

$$\frac{\partial\underline{\Delta}^*}{\partial b} = -(1 + vb)^{-2}v(vb + c - L - v) + (1 + vb)^{-1}v$$

$$= v(1 + vb)^{-2}(1 + vb - vb - c + L + v) = v(1 + vb)^{-2}(L + v - c + 1)$$

As shown in equation (20) that $L + v + 1 - c > 0$, $\frac{\partial\underline{\Delta}^*}{\partial b} > 0$.

Further, since $\frac{\partial b}{\partial x} < 0$, based on the chain rule of computing derivatives, we immediately have

$$\frac{\partial \underline{\Delta}^*}{\partial x} = \frac{\partial \underline{\Delta}^*}{\partial b}\frac{\partial b}{\partial x} < 0.$$

3. Proof for Lemma 2

Since $\frac{\partial \underline{\Delta}^*}{\partial x} = \frac{\partial \underline{\Delta}^*}{\partial b}\frac{\partial b}{\partial x} = \frac{\partial b}{\partial x}v(1+vb)^{-2}(L+v-c+1)$,

$$\frac{\partial^2 \underline{\Delta}^*}{\partial x \partial v} = \frac{\partial b}{\partial x}[-2(1+vb)^{-3}b(vL+v^2-vc+v)+(1+vb)^{-2}(L+2v-c+1)]$$

$$= \frac{\partial b}{\partial x}(1+vb)^{-3}[(1+vb)(L+2v-c+1)-2b(vL+v^2-vc+v)]$$

That is, to show $\frac{\partial^2 \underline{\Delta}^*}{\partial x \partial v} < 0$, because $\frac{\partial b}{\partial x} < 0$, it remains to prove

$(1+vb)(L+2v-c+1)-2b(vL+v^2-vc+v) > 0$

That is, to prove $L+2v-c+1+vbL+2bv^2-vbc+vb-2bvL-2bv^2+2bvc-2bv > 0$

That is, to prove $L+1+v+v+bvc > c+bv+bvL$

Since it has been shown $L+1+v > c$, it remains to show $v+bvc > bv+bvL$

First, since b<1, $v > bv$. Second, in equation (19), it needs to satisfy $vb+c-L-v > 0$. Since $vb < v$, c needs to be larger than L, and therefore $bvc > bvL$. So $v+bvc > bv+bvL$

4. Proof for Proposition 2

From the Lemma 2, it has been shown $\frac{\partial^2 \underline{\Delta}^*}{\partial x \partial v} < 0$. Thus, $\frac{\partial^2 \underline{\Delta}^*}{\partial x \partial \theta} = \frac{\partial^2 \underline{\Delta}^*}{\partial x \partial v} \cdot \frac{\partial v}{\partial \theta} > 0$, as $\frac{\partial v}{\partial \theta} = -V(1-\frac{1}{n}) < 0$.

$\frac{\partial^2 \underline{\Delta}^*}{\partial x \partial \theta} > 0$ suggests that the impact of the OSS patent pool on the threshold for new OSS product entry $\underline{\Delta}^*$ (OSS patent pool size x reduces this threshold) will be *lower* when θ is *high*. Therefore, on the other hand, the impact of the OSS patent pool on the threshold for new OSS product entry $\underline{\Delta}^*$ will be *higher* when the cumulativeness of innovation is high (i.e. when θ is *low*).

5. Proof for Proposition 3

From the Lemma 2, it has been shown $\frac{\partial^2 \underline{\Delta}^*}{\partial x \partial v} < 0$. Thus, $\frac{\partial^2 \underline{\Delta}^*}{\partial x \partial n} = \frac{\partial^2 \underline{\Delta}^*}{\partial x \partial v} \cdot \frac{\partial v}{\partial n} > 0$, as $\frac{\partial v}{\partial n} = -V\theta n^{-2} < 0$.

$\frac{\partial^2 \underline{\Delta}^*}{\partial x \partial n} > 0$ suggests that the impact of the OSS patent pool on the threshold for new OSS product entry $\underline{\Delta}^*$ (OSS patent pool size x reduces this threshold) will be *lower* when n is *high*. Therefore, on the other hand, the impact of the OSS patent pool on the threshold for new OSS product entry $\underline{\Delta}^*$ will be *higher* when the concentration of patent ownership is high (i.e. when n is *low*).

APPENDIX B: Patent Pledging Events

Event Date	Pledging Firm(s)	Pledged Patent(s)	Potential licensees	Notes
Jan 2005	IBM	More than 500 specified patents (contributed to the Patent Commons)	Anyone developing code under an OSI approved license	This pledging event has been included in our analysis
Jan 2005	Sun Microsystems	1670 (unspecified) patents related to Sun's Solaris	Developer working on any approved project under the Common Development and Distribution License (CDDL)	Two main criticisms of this pledge: 1. The CDDL doesn't permit mingling its code with code under GNU GPL, which governs Linux. This means developers can't use these patents on Linux – the freely granted patents can only enable idea-sharing among programmers for Solaris-related projects. 2. Sun's announcement was too broad and didn't specify these 1670 patents or respond to any developers' questions about what rights the developers have to these patents.
Sep 2005	Computer Associates International Inc.	14 patents (contributed to the Patent Commons)	Anyone developing code under an OSI approved license	This pledging event has been included as a control in our analysis.
Nov 2005	Nokia	Any of its patents	Developers working for the Linux Kernel only	Criticism: Because of Nokia's stance on Linux only, developers questioned why it did not apply to directly related projects such as GNOME and KDE and why it did not apply to application projects that are not necessarily directly related to Linux.
Nov 2005	Open Invention Network, founded by IBM, Novell, Koninklijke Philips Electronics, Sony and Red Hat	Any of OIN's patents	Any company, institution or individual that agrees not to assert its patents against the Linux operating system or certain Linux-related applications	This pledging event has been included in our analysis as a control, as many of its patents have been pledged only recently and toward the end of our sample period.
Feb 2007	Blackboard Inc.	Patent 6,988,138, 7,493,396, 7,558,853; pending patent applications: 12/470,739; 10/443,149; 10/643,075; 10/653,074; 11/142,965; 10/373,924; 10/918,016.	Anyone contributing to OSS projects, OSS initiatives, commercially developed open source add-on applications to proprietary products	For the commercially developed open source add-on applications to proprietary products, if the software's end license is open, then it is covered by the pledge; if it is partly open and partly proprietary, it is not covered.

APPENDIX C: Identification of Software Segments and the Matching Patent Classes

Step 1: Identify Software Segments

To measure entry with new OSS products related to different software segments in each year, a crucial step is to divide the software market into different segments that are reasonably distinct from each other. One main source of software segments is the product code classification system embedded in the PROMT database. For a portion of news articles from PROMT, there are a few product codes assigned to each new article that indicate what product category/categories are associated with that article. All these product categories are organized as a hierarchical structure by PROMT and are defined both in terms of customer segments and technologies. Table A-1 shows some examples of PROMT codes.

However, there are two drawbacks to just relying on PROMT classifications. First, a significant percentage (about 60%) of OSS product introduction news articles from PROMT is missing the product code field. Thus, we must manually assign product codes for this set of articles. Second, the PROMT classes do not include keywords, making it difficult to manually match articles to PROMT classes. Thus, we further match PROMT product code classes with CorpTech product code classes[43] to take advantage of the keywords defined for each CorpTech product code. The resulting concordance table (denoted as the PROMT-CorpTech concordance hereafter) consists of about 80 PROMT codes matched to CorpTech's six-digit or seven-digit product codes. Each product code is associated with a set of technology phrases specific to that product code. This is used as a basis for us to identify (i) the PROMT articles with missing product codes and (ii) the related patents across a variety of software segments. Table A-2 shows some examples of the PROMT-CorpTech concordance.

Step 2: Identify Patent Classes across Software Segments

Using the NBER patent data project and USPTO database, we constructed our patent dataset, which consists of all patents granted from 1976 to 2009. Our sample period is from 1999 to 2009. To identify the related patents across a range of market segments from the PROMT-CorpTech concordance, we first examined specialist firms that produce in only one software segment and particularly only one CorpTech six- or seven-digit code[44]. The sample of single specialists is from the CorpTech directory, over 1992 to 2004 and 2010.[45] We found 3500 patents held by about 700 specialists that operate in different software markets from the PROMT-CorpTech concordance. The 3-digit USPTO classes to which the 3500 patents and their forward citations belong served as a starting point for us to map patent classes to each product code: for each product code, the top decile of these 3-digit US classes was used as candidates representing

---

[43] 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.

[44] Examples of CorpTech code are provided in Table A-2.

[45] Unfortunately, data from the CorpTech Directory from 2005 to 2009 was not available.

the core technologies for that code. While the procedure we use is similar to the one used by Cockburn and MacGarvie (CM) (2011), we constructed our own classification for several reasons. First, our sample period is more recent than theirs, so the mapping between patent technologies and product markets may have changed over time. Second, Cockburn and MacGarvie examined 25 specific product codes that have incomplete overlap with the open source product market segments that we study. However, we did find many similarities between their classification and ours: for the product codes in both their and our classifications, the corresponding patent classes are very similar. Finally, we took the intersection of the patent classes from the patents in the OSS patent pool with the above mapping, which lead to 34 US patent classes and their corresponding product codes.

Step 3: Match Software Segments with Patent Class-subclass Combinations

Because most of the 3-digit US patent classes contain quite heterogeneous technologies, we then further generated a more detailed mapping between software product codes and US patent subclass levels by searching for technology phrases associated with each product code. This process generated the final mapping between software segments and patent class-subclass combinations. We further consolidated all product codes into 33 software segments based on whether they are supported by the same technologies (similar patent classes), as we are most interested in whether the supply of certain technologies by the OSS patent pools helps start-ups move into new technology area. The final concordance that we used in the empirical analyses consists of 33 software segments matched to 422 patent class-subclass combinations. Table A-3 shows some examples of this final concordance between software segments and US patent class-subclass combinations. Figures A-1 and A-2 present a more concrete view on the above three steps.

Keywords used to identify OSS entry

We used the following set of keywords to search in PROMT news articles for introduction of software products that are licensed as open source. A software product is tagged as open source if it contains any of these keywords. We first implement automatic search and then manually check the results to ensure it is licensed as open source. Our choice of open source license terms is based on the distribution of open source licenses used by OSS projects at SourceForge.net, which is the largest repository of OSS. Over 230,000 projects and over 3 million users and developers were registered before the end of year 2009 (SourceForge 2009).

Keywords related to generic terms of OSS:	open source , open-sourced, OSS, FLOSS, source code, GPL-compatible, non-copyleft, copyleft, free software license, open source license, open-source license, public domain
Keywords related to open source licenses:	GPL, General Public License , GNU, Lesser General Public License, LGPL, BSD, FreeBSD, Apache License, Apache Software License, Artistic License, MIT License, Mozilla Public License

Table A-1: Examples of PROMT 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
7372520	Peripheral support software
7372521	Device driver software
7372522	Data acquisition software
7372523	Printer support software
7372530	Disk/file management software

Table A-2: Examples of the PROMT-CorpTech Concordance

CorpTech Code	PROMT Product Code
SOF-CS-F	7372650 Fax software
SOF-DM-M	7372421 DBMS
SOF-HL-M	7372466 Medical practice software
SOF-ME-S	7372544 Sound/audio software
SOF-OA-MB	7372662 BBS software
SOF-OA-MC	7372674 Videoconferencing software
SOF-OA-ME	7372605 Electronic mail software
SOF-OA-MG	7372630 Workgroup software
SOF-OA-P	7372441 DTP software
SOF-TS-EC	7372433 Civil engineering software
SOF-TS-EE	7372434 Electrical engineering software
SOF-TS-ER	7372423 Geographic information systems
SOF-UT-H	7372521 Device driver software
SOF-UT-O	7372561 Data center management software
SOF-UT-Q	7372513 Application development software
SOF-UT-X	7372691 Data encryption software

Table A-3: Examples of the Concordance between Software Segments and US Patent Class-subclass Combinations

Software Segment	US class	Subclass Level 0	Subclass Level 1
Artificial Intelligence Software	706	Fuzzy Logic Hardware	Fuzzy Neural Network
Artificial Intelligence Software	706	Knowledge Processing System	Creation Or Modification
Artificial Intelligence Software	706	Knowledge Processing System	Knowledge Representation And Reasoning Technique
Artificial Intelligence Software	706	Neural Network	Learning Method
Artificial Intelligence Software	706	Neural Network	Learning Task
Artificial Intelligence Software	706	Neural Network	Neural Simulation Environment
Artificial Intelligence Software	706	Neural Network	Structure
Artificial Intelligence Software	706	Plural Processing Systems	
Data Encryption Software	380	Communication System Using Cryptography	Having Compression
Data Encryption Software	380	Communication System Using Cryptography	Time Segment Interchange
Data Encryption Software	380	Facsimile Cryptography	Including Generation Of An Associated Coded Record
Data Encryption Software	380	Key Management	Having Particular Key Generator
Data Encryption Software	380	Key Management	Key Distribution
Data Encryption Software	380	Particular Algorithmic Function Encoding	NBS/DES Algorithm
Data Encryption Software	380	Particular Algorithmic Function Encoding	Public Key
Data Encryption Software	380	Video Cryptography	Copy Protection Or Prevention
Data Encryption Software	726	Access Control Or Authentication	Network
Data Encryption Software	726	Access Control Or Authentication	Stand-Alone
Data Encryption Software	726	Monitoring Or Scanning Of Software Or Data Including Attack Prevention	Intrusion Detection
Data Encryption Software	726	Protection Of Hardware	Theft Prevention

Note: 1) US patent class 706 is described as "Data processing: artificial intelligence"; US patent class 380 is described as "Cryptography"; US patent class 726 is described as "Information security". 2) All subclasses within each US patent class are structured hierarchically. "Subclass level 0" means the subclass is on the highest level and "Subclass level 1" means the subclass is on the second highest level. Our mapping is based on subclass level 1.

Figure A-1: Identification of Software Segments

Software Segment	PROMT Product Code	CorpTech Product Code
Database software	(NT5) 7372420 Database software (NT6) 7372421 DBMS (NT6) 7372422 DBMS utilities	SOF-DM (Database/file mgmt. software) Keywords: Database/file management software, DBMS, Relational DBMS, Information storage and retrieval systems software (ISRS)
Image analysis software	(NT5) 7372450 Image processing software (NT6) 7372459 Image processing software NEC	SOF-OA-GI (Image processing software) Keywords: Image processing software, Image analysis software, Image enhancement software
Manufacturing and business process software	(NT6) 7372414 Business information management software (NT6) 7372416 Manufacturing, distribution and retailing software	SOF-MA(Manufacturing software) Keywords: Manufacturing automation protocol software, Operations planning software, Manufacturing planning software, Process control manufacturing systems software, Software to control product quality, Production scheduling software
Software development tools	(NT5) 7372510 Software development tools (NT6) 7372511 CASE software (NT6) 7372512 Programming utilities (NT6) 7372513 Application development software (NT6) 7372514 Debugging & testing software	SOF-PD (Program development soft.) Keywords: Software development systems, Development environment sof, IDEs, Language compilers, Program translator, program translators, Cross assemblers SOF-UT-C (Debugging and testing soft.) Keywords: Debugging and testing software

Step1

Figure A-2: Mapping Software Segments to Patent Subclasses

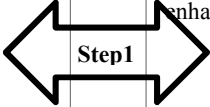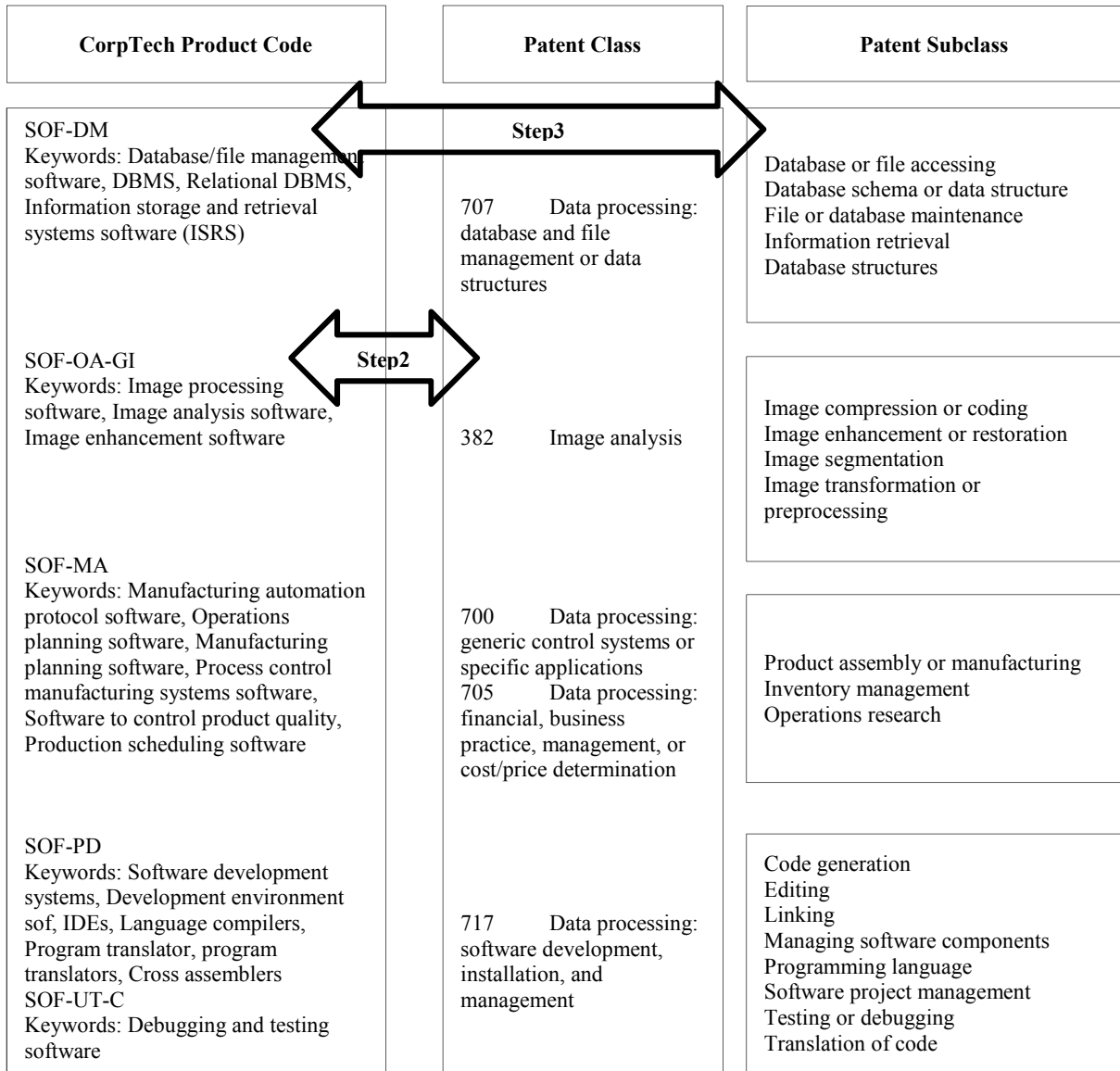| CorpTech Product Code	Patent Class	Patent Subclass
SOF-DM Keywords: Database/file management software, DBMS, Relational DBMS, Information storage and retrieval systems software (ISRS)	**Step3** 707 Data processing: database and file management or data structures	Database or file accessing Database schema or data structure File or database maintenance Information retrieval Database structures
SOF-OA-GI Keywords: Image processing software, Image analysis software, Image enhancement software	**Step2** 382 Image analysis	Image compression or coding Image enhancement or restoration Image segmentation Image transformation or preprocessing
SOF-MA Keywords: Manufacturing automation protocol software, Operations planning software, Manufacturing planning software, Process control manufacturing systems software, Software to control product quality, Production scheduling software	700 Data processing: generic control systems or specific applications 705 Data processing: financial, business practice, management, or cost/price determination	Product assembly or manufacturing Inventory management Operations research
SOF-PD Keywords: Software development systems, Development environment sof, IDEs, Language compilers, Program translator, program translators, Cross assemblers SOF-UT-C Keywords: Debugging and testing software	717 Data processing: software development, installation, and management	Code generation Editing Linking Managing software components Programming language Software project management Testing or debugging Translation of code

APPENDIX D: Supporting Empirical Results

Table A-4. C-test for GMM Estimates and

First Stage Results of OLS IV Regressions for the Direct Impact of an OSS Patent Pool

	(1)	(2)	(3)
GMM Estimates			
C-test statistic (p-value)	10.032 (0.123)	3.047 (0.803)	5.015 (0.542)
First Stage Results of OLS IV Regressions			
Dependent variable: OSS patent pool			
IBM patents*year2003_2004	.010 (.099)	.090 (.100)	.090 (.095)
square of IBM patents*year2003_2004	-.000 (.011)	-.010 (.011)	-.011 (.011)
IBM patents*afteryear2005	-.189** (.086)	-.002 (.100)	-.004 (.097)
square of IBM patents*after year2005	.086** (.010)	.066*** (.010)	.064*** (.011)
IBM opposed patents at EPO	1.886*** (.636)	2.328*** (.765)	2.375*** (.784)
square of IBM opposed patents at EPO	-.544*** (.169)	-.671*** (.191)	-.669*** (.190)
First Stage F-statistic (p-value)	83.63 (.00)	76.40 (.00)	35.17 (.00)
Controls	*Sales*	*Sales, SegmentPatents*	*Sales, SegmentPatents, PatentThicket*

Notes: 1) The full set of IVs include: IBM patents*year2003_2004, IBM patents*afteryear2005, square of IBM patents*year2003_2004, square of IBM patents*afteryear2005, IBM opposed patents in EPO, and square of IBM opposed patents in EPO. 2) The C-test is to assess the exogeneity of IBM patents*year2003_2004, IBM patents*afteryear2005, square of IBM patents*year2003_2004, and square of IBM patents*afteryear2005 as IVs with the null hypothesis that they are valid instruments. 3) The C-test statistic is computed as the difference between two J statistics from GMM estimates: that for the (restricted, fully efficient) regression using the full set of IVs versus that for the (unrestricted, inefficient but consistent) regression using the smaller set of IVs including IBM opposed patents in EPO and square of IBM opposed patents in EPO. 4) The first stage OLS IV Regressions are used as auxiliary regressions to test for weak IVs, as there is no such test in using the GMM estimator. 5) Robust standard errors, clustered by market segment, are in parentheses. 6) * significant at 10%, ** significant at 5%, *** significant at 1%.

Table A-5: Falsification Test: Direct Impact of an OSS Patent Pool

on Proprietary Software Products

We use the press releases of the 2,054 firms in the PROMT database to identify new proprietary software product entry. To identify products related to each segment, we focus only on introduction events associated with PROMT product codes. For each start-up, we include only the firm's first product in a segment to capture entry. This results in 2,384 proprietary product entry events from 2002 to 2009. We then aggregated these by software segment and year. We also adjust for a change in the assignment of product codes during our sample. Specifically, between 2007 and 2009 we found that application-related software products were systematically assigned to a higher product code level in the PROMT database (specifically, they were assigned to 7372400, Applications Software). This forced us to combine several application segments together, leaving us with 29 software segments in total.

Dependent variable	New proprietary software product entry		
	(1)	(2)	(3)
OSS patent pool	.007 (.122)	-.011 (.059)	.003 (.065)
Sales	.793 (.505)*	.470 (.530)	.364 (.516)
Total patents		6.841** (2.736)	4.149 (3.095)
Patent quality		7.037* (4.033)	3.793 (4.606)
OIN patents			.034 (.049)
SSO patents			-.116** (.052)
Observations	232	232	232
Log pseudolikelihood	-496.316	-464.505	-451.968

Notes: 1) Robust standard errors, clustered by market segment, are in parentheses. 2) * significant at 10%, ** significant at 5%, *** significant at 1%. 3) All regressions include market and year fixed effects.

Table A-6: GMM Estimates of Count Data Regression Using Different Sets of IVs

Dependent variable: OSS entry					
	Specification without interactions		Specification with two-way interactions		
	(1)	(2)	(3)	(4)	(5)
OSS patent pool	.432*** (.163)	.290*** (.133)	-.250 (.295)	-.518** (.217)	-.371 (.321)
cumulativeness			.472 (2.126)	-.315 (2.199)	.479 (2.164)
OSS patent pool * cumulativeness			.361 (.315)	.396* (.261)	.479* (.276)
Concentration			-5.334 (6.275)	-8.291 (6.666)	-7.365 (6.522)
OSS patent pool * Concentration			1.655** (.705)	2.027*** (.564)	1.451*** (.898)
Potential licensors			.779 (1.180)	-1.727 (1.505)	-.788 (2.004)
Sales	-.091 (.317)	-.223 (.204)	-.291 (.255)	-.280 (.225)	-.319 (.227)
Patent quality		1.044 (.675)		2.438 (1.678)	2.322 (1.674)
Total patents				2.625 (1.738)	2.021 (1.745)
SSO patents					-.111 (.129)
Seven year dummies	Yes	Yes	Yes	Yes	Yes
Observations	363	363	363	363	363
Over-identification test statistic (p-value)	8.914 (0.11)	15.691 (0.11)	16.810 (0.10)	17.522 (0.13)	19.096 (0.04)

Notes: 1) The full set of IVs include: IBM patents*year2003, IBM patents*afteryear2004, square of IBM patents*year2003, square of IBM patents*afteryear2004, IBM opposed patents at EPO, square of IBM opposed patents at EPO. 2) Robustness standard errors, clustered by market segment, are in parentheses. 3) * significant at 10%, ** significant at 5%, *** significant at 1%. 4) Adding more controls in specification without interactions leads to non-convergence of the GMM estimator.

Table A-7. Conditional Fixed-effect Poisson Regression: Direct Impact of an OSS Patent Pool and its Interaction with Cumulativeness of Innovation (using robust measure of cumulativeness)

Dependent variable: OSS entry	Robust Measure of Cumulativeness		
	(1)	(2)	(3)
OSS patent pool	.073 (.123)	.054 (.111)	.093 (.142)
Cumulativeness	1.119* (.623)	1.840*** (.697)	2.464*** (.943)
OSS patent pool * Cumulativeness	.406** (.176)	.458*** (.176)	.549*** (.166)
Sales	-.129 (.519)	-.067 (.521)	-.055 (.496)
Potential licensors		-.201 (1.635)	.059 (1.634)
Total patents		.178 (2.247)	-.585 (2.462)
Patent quality		-1.322 (2.353)	-1.781 (2.430)
Concentration			-14.924** (6.438)
OIN patents			-.057 (.067)
SSO patents			-.016 (.065)
Observations	286	286	286
Log pseudolikelihood	-226.467	-225.946	-223.847
Marginal Effects			
OSS patent pool (average)	.238* (.129)	.239** (.125)	.315** (.147)
OSS patent pool (cumulativeness=10%)	.108 (.121)	.093 (.109)	.140 (.140)
OSS patent pool (cumulativeness=90%)	.482** (.193)	.515*** (.188)	.648*** (.202)
Statistic for the test of the difference (p-value)	5.32** (0.02)	6.80*** (0.01)	10.96*** (0.00)

Notes: 1) Robust standard errors, clustered by market segment, are in parentheses. 2) * significant at 10%, ** significant at 5%, *** significant at 1%. 3) The statistic for the test of the difference of marginal effect of an OSS patent pool between the cumulativeness of innovation at the 10th percentile and at the 90th percentile is distributed as chi-square, and the p-value is shown in the parentheses. 4) All regressions include market and year fixed effects.

Table A-8. Conditional Fixed-effect Poisson Regression: Direct Impact of an OSS Patent Pool, its Interactions with Cumulativeness of innovation and with Patent Ownership Concentration, and Three-way Interaction

Dependent variable: OSS entry			
	(1)	(2)	(3)
OSS patent pool	.232 (.254)	.330 (.264)	.389 (.277)
Cumulativeness	2.741 (2.004)	3.268 (2.280)	3.183 (1.954)
OSS patent pool * Cumulativeness	-.346 (.385)	-.434 (.403)	-.490 (.383)
Concentration	-11.076* (6.869)	-8.843 (6.959)	-10.113 (7.445)
OSS patent pool * Concentration	-1.048 (.831)	-1.360 (.884)	-1.483 (.916)
Cumulativeness *Concentration	-.256 (5.351)	-1.450 (6.006)	-.664 (5.306)
OSS patent pool * cumulativeness *Concentration	2.668* (1.502)	3.096** (1.618)	3.318** (1.562)
Potential licensors	.860 (.615)	.001 (1.586)	-.121 (1.516)
Sales	.116 (.518)	.106 (.514)	.115 (.493)
Total patents		-.138 (2.468)	.220 (2.525)
Patent quality		-2.404 (2.658)	-2.464 (2.642)
OIN patents			-.074 (.068)
SSO patents			-.021 (.069)
Observations	286	286	286
Log pseudolikelihood	-222.779	-222.294	-221.829
Marginal Effects			
OSS patent pool (average)	.182* (.103)	.216** (.106)	.242** (.116)
[1] OSS patent pool (cumulativeness=10%, concentration =10%)	.095 (.141)	.146 (.143)	.183 (.155)
[2] OSS patent pool (cumulativeness=10%, concentration =90%)	.001 (.097)	.008 (.097)	.028 (.097)
[3] OSS patent pool (cumulativeness=90%, concentration =10%)	.097 (.252)	.104 (.255)	.104 (.255)
[4] OSS patent pool (cumulativeness=90%, concentration =90%)	.749*** (.169)	.830*** (.199)	.877*** (.200)
Statistic for the test of the difference between [1] and [4] (p-value)	18.96*** (0.00)	16.38*** (0.00)	16.90*** (0.00)
Statistic for the test of the difference between [2] and [4] (p-value)	14.28*** (0.00)	12.88*** (0.00)	13.60*** (0.00)
Statistic for the test of the difference between [3] and [4] (p-value)	3.44* (0.06)	3.76** (0.05)	4.70** (0.03)

Notes: 1) Robust standard errors, clustered by market segment, are in parentheses. 2) * significant at 10%, ** significant at 5%, *** significant at 1%. 3) The above estimation is based on the specification $X_{jt}'\beta = \beta_1 OSS\ patent\ pool_{jt} + \beta_2 OSS\ patent\ pool_{jt}*cumulativeness_{jt-1} + \beta_3 OSS\ patent\ pool_{jt}* concentration_{jt-1} + \delta_1 OSS\ patent\ pool_{jt}*cumulativeness_{jt-1}*concentration_{jt-1} + \delta_2 cumulativeness_{jt-1}*concentration_{jt-1} + \gamma_1 Sales_{jt} + \gamma_2 SegmentPatents_{jt-1} + \gamma_3 PatentThicket_{jt-1} + \gamma_4 OtherFreePatents_{jt} + \tau_t$. 4) The difference between [1] and [2] / between [1] and [3] / between [2] and [3] is insignificant and not included in the table due to the limited space. 5) All regressions include market and year fixed effects.