Patent Commons, Thickets, and Open Source Software Entry by Start-Up Firms*

Wen Wen
McCombs School of Business
The University of Texas at Austin
1 University Station, B6000
Austin, TX 78712
wen.wen@mccombs.utexas.edu

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

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

July 2013

Abstract

We examine whether the introduction of a patent commons, a special type of royalty free patent pool available to the open source software (OSS) community influences new OSS product entry by start-up software firms. In particular, we analyze the impact of The Commons—established by Open Source Development Labs and IBM in 2005. We find that increases in the size of The Commons related to a software market increase the rate of entry in the market by start-ups using an OSS license. The marginal impact of The Commons on OSS entry is increasing in the cumulativeness of innovation in the market and the extent to which patent ownership in the market is concentrated.

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

JEL Classification: O34, L86

* We thank Ajay Agrawal, Tim Bresnahan, Alberto Galasso, Stuart Graham, Josh Lerner, Megan MacGarvie, Timothy Simcoe, and participants at the NBER Patents, Standards, and Innovation Pre-Conference, the 2011 Workshop on Information Systems and Economics (WISE), the 2011 Conference on Information Systems and Technology (CIST), and the 2011 REER conference at the Georgia Institute of Technology. We gratefully acknowledge funding from a Kauffman Foundation/Georgia Research Alliance grant for the study of entrepreneurship and productivity. We thank Nirmalya Choudhury, Matthew Espy, Emily Getreu, Bridget M. Gorta, Sujay Mehta, Daniel Mitchell, and Jian Zhao for outstanding research assistance. Wen Wen thanks the Kauffman Foundation for providing funding for this research through a Kauffman Dissertation Fellowship. Chris Forman acknowledges funding from the Alfred P. Sloan Foundation through an Industry Studies Fellowship. All errors are our own.

1. Introduction

While patents play an important "property rights" role in facilitating transactions in markets for technology (e.g., Arora et al. 2001; Gans et al. 2002; Teece 1986), the strategic use of patents and the appearance of dense, overlapping webs of property rights known as patent thickets may also work to stifle innovation (e.g., Bessen and Meurer 2008; Jaffe and Lerner 2004; Shapiro 2001). Empirical evidence on this issue is still mixed, with implications varying across industries and firms.[1] The presence of patent thickets may be a particular issue for developers of information technology (IT) products like computing hardware and software, where innovation is often highly cumulative and products rely on standards of heterogeneous inventions for whose patent rights are often owned by many different firms.[2] Small firms will often find navigating such patent thickets particularly difficult, as they will typically not have a large intellectual property portfolio that they can use to cross-license with existing IP holders.

A range of institutional mechanisms have appeared to help companies navigate patent thickets, including standard setting organizations (SSOs) and patent pools. Some of these mechanisms, such as traditional patent pools and SSOs that make patents available under "reasonable and non-discriminatory terms, lower the transaction costs of identifying and negotiating licensing agreements for related technologies, but may also increase the incentives for some patentees to litigate (e.g., Lampe and Moser 2013; Simcoe, Graham, and Feldman 2009) As a result, their effects on the use of cumulative inventive activity building upon covered technologies have been mixed.[3]

An alternative approach is to form institutions under which IPR are made available royalty free, requiring neither a licensing fee nor grant-backs to the community. One recent development has been to offer IPR through a patent commons, a special type of patent pool that offers royalty-free usage of patents, generally covering a broad set of technologies, to any firm that promises not to assert their IPR against the commons' beneficiaries (Lévêque and Ménière 2007, Serafino 2007, Hall and Helmers 2011). While the number and size of such donations have been growing rapidly, thus far we have little evidence on whether they in fact contribute to increased innovative activity.[4] This is an important gap in understanding. The

[1] For example, Hall and Ziedonis (2001) study on the semiconductor industry finds that the pro-patent shift in U.S. policy in the 1980s spawned an increase in strategic patenting among capital-intensive firms, but also facilitated entry by specialized design firms. Cockburn and MacGarvie (2011) find that, following expansions in the patentability of software in the mid-1990s, average entry rates declined but the likelihood of entry by firms holding patents increased.

[2] For example, Biddle, White, and Woods (2012) identify 251 technical interoperability standards in a modern laptop. Of these 251 standards, 112 were developed by consortia, 90 by formal standards development organizations, and 49 were developed by individual companies.

[3] For example, Rysman and Simcoe (2008) show that citations to patents increase significantly after disclosure that they are part of a standard, while Lampe and Moser (2010) study the sewing machine patent pool and find that the pool decreased patenting and innovation, particularly among members of the pool.

[4] One recent exception is Hall and Helmers (Forthcoming), whose recent study of the "Eco-Patent Commons" showed that it had "no discernible impact on the diffusion of the knowledge embedded in the protected technologies to other patenting firms" (Hall and Helmers 2011). However, their results must be interpreted with caution, given the short period of time following the establishment of this particular patent commons (2008).

literature suggests that inventors may have insufficient incentives to provide intellectual property as a public good to the community (Gambardella and Hall 2006), and so the patent commons may contain few or low-quality patents that have little impact on innovative activity.[5] In addition to the potential low quality issue associated with poor incentives highlighted earlier, the patents pledged in the commons may not contain all of the technologies necessary to produce and commercialize a new invention. If other necessary technologies are not held within the commons, barriers to entry may still be high. In short, our research speaks to the efficacy of a growing mechanism to help firms navigate patent thickets.

Motivated by these observations, we take a first step to evaluate how the creation of a patent commons affects innovation in the open source software (OSS) setting. Our key argument is that because patent commons could potentially mitigate licensing cost and litigation threats caused by patent thickets, start-ups that have access to the commons would face lower sunk costs of entry than otherwise and thus be more likely to enter into downstream markets. Using a similar logic, we argue that such institutional mechanisms should be particularly valuable in environments with high licensing and litigation costs. We follow the existing literature and highlight the following two settings where such costs would be the highest and thus patent commons would have the greatest impact on OSS entry—when innovations are highly cumulative and when IPR ownership (particularly patent ownership) is very concentrated.We examine how the introduction of a patent commons, initiated by the Open Source Development Labs and IBM in 2005 (commonly referred to as "The Commons") influences entry by startups using an OSS license (which we refer to as "OSS entry").[6] We assemble data on OSS entry using longitudinal press releases from 2,054 U.S. start-up software firms contained in the Gale database "PROMT". Using count data conditional fixed effects models, our empirical strategy examines whether time series variation in the number of patents contributed to The Commons related to a narrowly defined software market is associated with changes in the number of OSS entrants into that market. A key and challenging task in assembling our dataset was to identify patents relevant to innovation and entry in product markets. 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 markets based upon the technological classes of patents and their key words.

Our initial econometric approach assumes that changes in the number of The Commons' patents are uncorrelated with omitted variables that may influence OSS entry by start-ups. Our results show that a 10% increase in The Commons' patent claims in a software market is associated with a 1.3%-2.9%

[5] While we do not explicitly focus our research on the question of why firms may contribute some of their patents to the commons, we do discuss IBM's incentives for establishing a patent common in 2005 to motivate our analysis as well as to probe our identification strategy (see sections 3.2 and 6.3).
[6] 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.

increase in the rate of OSS entry by start-ups into that market.[7] However, introduction of The Commons influences entry only in those markets where the ex ante costs of litigation and licensing are highest, namely those where innovations are highly cumulative and when IPR ownership is concentrated. Figures 2, 3 and 4, which show how OSS entry is shaped by the introduction of The Commons and the extent of cumulativeness and concentration in the market, forecast our main results. Figure 2 shows that in markets with a large number of The Commons' patents there is a significant entry after introduction of The Commons, while there is no change in entry in other markets. Figures 3a and 4a show that there is little growth in entry after the introduction of The Commons in markets where The Commons has few patents. Figures 3b and 4b examine growth in entry in markets where The Commons is well represented; Figure 3b shows that the introduction of The Commons is associated with greater entry but only in markets where innovation is cumulative, while Figure 4b shows that growth in entry post-Commons is higher in markets with high IPR concentration.

A particular concern is that donations to The Commons are the result of decisions by mostly one firm, IBM. If there are changes in unobserved factors that influence entry and the size of The Commons in a market, such as sales growth, then this could cause is to misinterpret our results as reflecting the causal benefits of The Commons. We examine concerns about how this potential alternative explanation, as well as other sources of omitted variable bias, might influence our results. First, in addition to our baseline results that control for several time-varying market features, we show that our results are robust to controlling for market-specific time trends. Second, we show that growth in The Commons is not associated with increased entry from products that should not see entry costs fall with its introduction, namely products offered under a proprietary license. Next, we examine the timing of the effects of The Commons. We find that growth in The Commons is not associated with increased entry between 1999-2003, but do find that it is associated with entry in 2004, one year prior to its introduction. We provide evidence that this finding is likely caused by other activities that IBM was engaged in to promote open source over the same period.

We estimate count data models with instrumental variables using the Generalized Method of Moments (GMM) to address the potential endogeneity of the patent commons variable. Our first instrument strategy uses variation in IBM patents litigated at the European Patent Office (EPO). Our identification strategy is based on the assumption that this variable is correlated with IBM's holdings of blocking patents across different narrowly defined software markets, which should influence its propensity to pledge. Our other instruments use an exogenous change in the perceived risks to OSS and IBM support to OSS that occurred during our sample period; namely, SCO's lawsuit against IBM for

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

copyright infringement in the Linux kernel (Alexy and Reitzig 2013). Our second instrument interacts our first with a time dummy that turns on after 2003 to capture how changing incentives to contribute to The Commons may interact with IBM's patent holdings. Finally, since this enforcement is particularly related to IBM's deployment of Linux operating system, our third instrument is the three-way interaction among the first, the time dummy that turns on after 2003, and a dummy that turns on for operating system software market. Our results using these models are consistent with the baseline described above. While we acknowledge that none of these analyses individually eliminates the possibility that our results are influenced by omitted variables, together they collectively circumscribe the manner in which omitted variables would need to influence our results.

Several considerations support the significance of our research question. First, despite the rapid growth in the commercialization of OSS,[8] OSS contributors and users are easily targeted by proprietary holders of IPR because of the cumulative nature of the development process and the difficulty in identifying the provenance of source code ownership. The existence of fear, uncertainty, and doubt (FUD) arising from the threat of litigation will be important for the commercialization of OSS. In general, while OSS emphasizes open access to the source code and avoids the use of formal appropriability mechanisms, there has been little understanding of how the existence and exercise of formal IPR influence OSS innovation (Lerner and Tirole 2005a, von Hippel and von Krogh 2003).

Related, our research has important implications for technology management. Contributors to patent commons forego potential opportunities to license their IPR in hopes of increasing innovative activity that will spur demand for complementary products and services from which the contributor can appropriate value. If contributions do not lead to entry of new products using the contributed technologies, then the benefits to contribution will be isolated to the positive publicity in the OSS community.

Last, our findings inform recent work on the impacts of patent thickets on start-up innovation (e.g., Cockburn and MacGarvie 2011). 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.[9] Start-ups may be offering complementary technologies within different platforms and may be unaware of the extent to which they are infringing patents or to which patents in The Commons offer a means to avoid patent infringement.

2. Theoretical framework

2.1 How patent commons lower the sunk costs of entry for OSS firms

[8] For example, 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).

[9] 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)

Patent thickets can increase at least three broad types of sunk costs of entry by start-ups: (1) the costs of inventing around existing patents; (2) the costs of infringement, which may include the costs of licensing the infringed technology and the costs of litigation such as acquiring a defensive patent portfolio as well as injunction and damages; and (3) the transaction costs of acquiring patents owned by others.

While firms that have a defensive patent portfolio may be able to navigate the patent thicket through cross-licensing their own patents with those of other IPR holders, this strategy will be harder to implement for OSS firms. OSS developers are often philosophically opposed to software patents, considering them antithetical to the spirit of freedom that imbues OSS development (Schultz and Urban 2012; Stallman 2011; Marson 2004).[10] Further, the costs of writing and administering patents may be too high for small, OSS developers or firms relative to their benefits. For example, small OSS firms may have insufficient resources to hire legal staff.

As has been noted elsewhere, firms with significant patent portfolios and who use traditional appropriability mechanisms to commercialize new technologies in other settings will sometimes also contribute to open source communities under a so-called "private-collective model of innovation" (von Hippel and von Krogh 2003). Under this model, firms may contribute to public goods such as OSS. While they may be unable to appropriate value directly from the public good, they may be able to create and appropriate value from complementary products and services such as downstream application software that interfaces with open source or support for open source software.[11] While such firms may be able to navigate patent thickets through the use of their own patent holdings in cross-licensing, the inability of other small firms to similarly navigate the thicket may reduce the extent of complementary innovation and in so doing reduce the value of the public good.

One institutional mechanism that existing patent holders have used to stimulate complementary innovation is to donate patents to the community. Such donations go beyond the requirement that contributors to OSS products automatically grant a license to use, modify, and redistribute the code to all other legitimate users of the code but also extends to complementary technologies that may be used in conjunction with existing OSS products. Such contributors not only incur the opportunity cost of foregone profits from not enforcing the IPR, but also bear an explicit cost of maintaining and renewing the patent. These contributions have often been labeled royalty-free patent pools, or patent commons (Hall and

[10] OSS firms may fear the existence of patents that are even pledged to the community for purely defensive purposes; while the initial motivates may be altruistic, it remains possible that the original patent holder may have a change in strategy, in particular a change in ownership or management may initiate a change in how patents are used within the organization (e.g., Schultz and Urban 2012).

[11] For example, Fosfuri et al. (2008) find that firms will be more likely to produce open source software products when they have large stocks of complementary patents or downstream complementary capabilities.

Helmers 2012). While the motivations to contribute to these patent commons are in themselves interesting, it is not our goal to explore them, as they have been investigated elsewhere (Alexy and Reitzig 2013).

Contributions to patent commons can have direct and indirect benefits to potential OSS producers. A first direct benefit is the reduction in expected invention and licensing costs related to the donated technology. Instead of inventing around, OSS firms can directly use the technology protected by the commons' patent. Moreover, start-ups could also possibly use the pledged patent to substitute blocking patents that are not in the commons, avoiding the costs of transacting and negotiating with patentees who have not contributed their IPRs. These benefits will be increasing in the number of patents that are made available to OSS producers. Second, as incumbents increase their contributions it will signal their intention to generate profits as a private-collective (p-c) innovator through the production of complementary goods and services, rather than through direct enforcement of IPRs. While the contributor retains the right to enforce other patents not included in the donation, doing so would harm its reputation among OSS contributors, decreasing its ability to create and appropriate value as a p-c innovator. As a result, the perceived threat of litigation will decline, or, to use a widespread terminology among software industry practitioners, there will be less "*fear, uncertainty, and doubt"* (FUD, cf. Auza 2011).

However, there may be indirect benefits as well. As Alexy and Reitzig (2013) describe, the donation of patents may encourage reciprocal behavior from other industry participants that may benefit p-c innovation. For example, other firms involved in private-collective innovation may also choose to pledge patents for use by the community, and may acquire additional patent rights to prevent proprietary innovators from acquiring the same and using them against the community. As more patents in a technology area are pledged in the commons, it is more likely to create norms of non-enforcement and encourage other private-collective innovators' contribution. One way to view this result is through the lens of public goods. The contribution of patents represents a commitment not to assert IPRs with the goal of fostering the development of new software development by entities who are unable to protect themselves against IPR enforcement by traditional means. This goal is more likely to be accomplished when contributions are made by multiple firms. However, in an asymmetric information environment, firms have an incentive to undercontribute (Coase 1960). To provide the public good, actors with high valuations must contribute more than those with lower valuations (Mailath and Postlewaite 1990; Al-Najjar and Smorodinsky 2000). In our setting, this corresponds to the initiator of the commons providing more than other firms, and so the likelihood of forming successful patent commons will be increasing in the initial contribution. In sum, we expect that patent commons may effectively reduce licensing costs and litigation threats faced by the start-ups, and such benefits will be increasing in the size and scope of the original pledge. It is also worth noting that these positive effects on entry would be reinforced as potential OSS producers enter and produce complementary OSS products, as these complementary products would create a platform of interlocking components, which increases the value of OSS products and services to potential buyers and thus further facilitate entry.

2.2 How the value of the commons varies with market characteristics

In the previous section we showed how a patent commons can reduce entry costs by reducing expected invention costs and by alleviating infringement costs and transaction costs of negotiating with patentees. In this section we explore market circumstances where the effects of such a patent commons on reducing entry costs, and so encouraging entry, will be greatest.

We focus on two observable dimensions of patent thickets that influence the ex ante costs of entry, where we should expect to see the effect on entry through introduction of the commons would vary, other things equal. The first is the cumulativeness of innovation, defined as the extent to which an innovator builds on prior developments and discoveries (Scotchmer 2004: 127). In environment with high cumulativeness, the boundaries of potential blocking patents usually blur, which makes it difficult to build upon existing patents, leading to high costs of inventing around. Second, highly cumulative innovations also suggests high infringement costs and transaction costs, as start-ups could easily infringe and may need to obtain licenses for a large set of related patents to enter into a technology space. These costs are particularly high when prior inventions are protected by broad patents (Merges and Nelson, 1990; Scotchmer 1991; Bessen and Maskin 2009). Therefore, because patent commons could reduce these costs and become most valuable when these ex ante costs are highest, it will have the greatest effect on entry in environments with cumulative innovations.

The second market characteristic that we study is the concentration of IPR ownership within a software market, defined as the extent to which IPR are distributed across patent holders. Two views have recently been set forth about how concentration influences the costs of licensing negotiations. One view holds that increases in fragmentation of IPR (i.e. decreases in concentration of IPR) will increase the transaction costs of licensing IPR, creating an "anti-commons" effect (Heller 1998, 1999; Heller and Eisenberg 1998). Under this view, when there are many small exclusionary IPR that are held by many firms, the costs of coming to terms with many IPR holders will influence a firm's strategic response to potential expropriation risks from external patent holders. For example, Ziedonis (2004) shows that firms will patent more aggressively when property rights are fragmented. Under this view, the costs of acquiring the rights to use inventions owned by others are particularly high when IPR are fragmented.

Recent work has challenged the anti-commons view. Under this view, concentration of IPR increases the value of the negotiation for the licensor, increasing the incentives to litigate (Lichtman 2006; Galasso and Schankerman 2010). For example, consider the case where property rights are alternatively held by 1000 firms or 1 firm. While transaction costs are potentially higher in the former setting, the potential licensee may not need to obtain licenses from each of the 1000 firms as some patentees may not litigate the patent if there are non-zero costs to litigating the patent and some uncertainty of whether the focal patent will be upheld. Galasso and Schankerman (2010) formalize this intuition, showing in the context of a bargaining game that fragmentation reduces the negotiation value of a patent that is potentially infringed and reduces the time to settlement in a patent dispute. In short, under this second view, the expected costs of infringement are highest when IPR are highly concentrated, and in such environment, an institution such as patent commons that could effectively reduce licensing costs and litigation threats would encourage entry.

3. Research Setting

3.1 The Commons.

We focus on one major patent commons—the Open Source Development Labs' Patent Commons project (commonly referred to as "The Commons"). In January 2005, IBM pledged access to 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 use, sponsor, or develop OSS pledged additional 29 patents to The Commons.[12] "Pledge" in this context means that "patent holders agree they will not, under certain terms and conditions, assert patent rights against third parties who are engaging in activities that might otherwise give rise to a claim of patent infringement."[13] IBM announced in its press release that it believed this was the largest patent pledge of any kind. All pledged patents are explicitly listed on an online public database, and users of the technologies embedded in the patents are not required to sign any formal agreement with The Commons.

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 A for a detailed summary of these events), our choice of The Commons as the focus of our analysis is guided by several factors. First, The Commons is economically important in the sense that it comprises a large collection of patents. Second, The Commons covers multiple software technology markets, allowing us to use variation over time within software markets for identification. Third, The Commons was introduced in 2005, allowing sufficient time to observe changes in entry behavior after its introduction. Fourth, The Commons specifies contributed patents at a very detailed level, listing each of their patent numbers. We could find no other patent donation that similarly met all of these criteria. Nevertheless, in our empirical analysis we control for the effects of other commons-like institutions.

3.2 IBM historical support for open source software

[12] Example companies include Computer Associates International Inc. and Open Invention Network, LLC.
[13] For more details, see http://www.patent-commons.org/resources/about_commitments.php.

The establishment of The Commons is consistent with IBM's strategy since the late 1990s of supporting OSS (Capeck et al. 2005; Samuelson 2006; Campbell-Kelly and Garcia-Swartz 2009). IBM initially announced its commitment to Linux in 1999, and in 2001 announced that it would invest $1 billion over the following three years to make Linux more suitable for enterprise applications (Campbell-Kelly and Garcia-Swartz 2009). In keeping with this commitment, IBM has made all of its hardware platforms compatible with Linux, released Linux versions of its software products, and developed Linux-focused service capabilities. IBM explicitly supports OSS to promote open standards in areas that are complementary to its profitable businesses (Capek et al. 2005). Over time, it has focused less on operating systems, and focused more on developing and marketing middleware or application software. IBM's business model now focuses on selling high end hardware, proprietary software running on top of Linux, and systems integration and other customized services to enterprise customers (Samuelson 2006).

IBM's contributions to create and license OSS throughout our sample period may cast doubt about the real effects The Commons may have on innovation by other OSS firms. As a practical matter, an OSS contributor implicitly grants a license to any patents that author holds that read on the software that has been developed; otherwise any users of that OSS may become inadvertent infringers of the patent. However, while some OSS licenses, such as the General Public License v3 (GPL v3), include an explicit grant of a patent license, many do not (Capek et al. 2005). Further, it may own IPRs that do not directly cover contributed technology but may be closely related to it. This contributes to create legal uncertainties for both producers and users of OSS.

In summary, our analysis of IBM's involvement with OSS suggests that its support for OSS starts before and continues throughout our sample period (1999 to 2009, as explained below). However, the establishment of The Commons represents a significant discontinuity in terms of IBM's *explicit* legal support for OSS. The only similar event that occurs during our sample period was IBM Senior Vice President Nick Donofrio's announcement during a keynote at the LinuxWorld conference that IBM would not assert its patents against the Linux kernel (Scannell 2004).[14] We will examine the effect of this 2004 announcement and its relationship with the establishment of the Commons in 2005 in our sensitivity analysis.

4. Data

4.1. Sample

Our sample consists of 2,054 US software firms from the 2004 and 2010 editions of the

[14] It is believed that this announcement was made in response to a finding by the Open Source Risk Management organization that several large companies, including IBM, held patents that might affect the Linux kernel (Scannell 2004).

CorpTech Directory of Technology Firms[15] (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 US-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 markets of firms, but have little time variation, while the NETS data have limited product market information but vary over time.

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.[16] Our sample period is from 1999 to 2009, with 6 years before the establishment of The Commons and 5 years after. We believe this time window is sufficiently long to capture the impact of The Commons on OSS entry.

4.2. Identifying software markets 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 markets.[17] Because of certain drawbacks of only relying on the PROMT classifications (we describe these in further detail in Appendix B), we further match PROMT's software-related product categories with CorpTech's "SOF" product classes[18] to create a PROMT-CorpTech concordance so that each PROMT software 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 markets with the most relevant patent classes as described below.

An important part of our data construction involves matching product markets to patents. This allows us to identify both the cumulativeness of innovation and the concentration of patent ownership in a software market. As is well-established in the literature, this type of matching is difficult (e.g., Griliches, 1990, Silverman 1999). 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 markets are classified through PROMT categories, in order to create the final mapping between software markets and patent classes, we then combine the PROMT-CorpTech concordance and patent-CorpTech

[15] Our choice of 2010 CorpTech data reflects a constraint with the data—we have contacted CorpTech and there are no historical data from 2005 to 2009, the core years of our sample period. The combination of CorpTech 2004 and 2010 is to address potential survivor bias.

[16] 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.

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

[18] There are more than 290 software product codes (denoted as SOF categories) 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. For further details on this classification system, see Cockburn and MacGarvie (2011).

concordance to form the PROMT-patent concordance. The final concordance that we use in the empirical analysis consists of 33 software markets matched to 422 patent class-subclass combinations[19] (see the Appendix B for a detailed discussion of our data construction process).

5 Measures

5.1. Dependent variable: OSS entry

This variable refers to the number of OSS entrants into software market *j* in year *t*. We use a three-step procedure to identify new OSS entry in a software market based on the press releases of the 2,054 firms in the PROMT database. First, following the 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 B 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)[20] 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 into a market by a start-upmarket. In total, we have 242 new OSS product entry events made by 85 start-up firms from 1999 to 2009.[21] We aggregated these new OSS product entries by software market and year. Our dependent variable is therefore equal to the number of new OSS start-up entrants in market *j* and year *t*. The data are structured as a balanced panel. Table 1 includes a brief description of measures and summary statistics for the main variables used in our empirical analysis.

5.2. Independent variables

The Commons. This variable is equal to the number of claims-weighted patents in The Commons related to software market *j* in year *t*. 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 patents in The Commons will be more likely to invalidate blocking patents. To capture these effects, we measure the claims-weighted count of patents in The Commons related to each software market.[22] We further take the logged value of this variable[23] to reduce skewness.

---

[19] Table B-3 in Appendix B lists examples of this final concordance between software markets and US patent class-subclass combinations.

[20] 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).

[21] 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.

[22] We also use raw patent counts as a robustness check. The results are qualitatively similar to the claims-weighted measure and are reported in Table C-2 and Table C-3 in the Appendix.

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

Cumulativeness. This variable refers to the cumulativeness of innovation within market *j* in year *t*. 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 market *j* and year *t* to backward cite patents within the same market *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 market *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 market) 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 market), (*n*-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 market *j* is based on the share of potential backward citations to patents belonging to the same market that are actually cited by the focal patent. The cumulativeness of innovation for software market *j* is then the average of all *N*-1 patents' cumulativeness:[24] $C_j = \frac{\sum_{n=2}^{N} \sum_{m=1}^{N} \frac{x_{nm}}{n-1}}{N-1}$. This measure varies over time based on the grant year of the market *j* patents under consideration. Notice that the oldest patents in a market 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 market. 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 market. 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 market *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 market *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

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

concentration ratio for market 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 market j up to year t.[25]

5.3. Control variables

Sales growth. One important factor that may correlate with both the behavior of firms contributing to The Commons and OSS entry by start-ups is the rate of market growth for software market j, which is proxied by the sales change from year t-1 to year t in market j. 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 yearly sales change for each of the eight SIC categories and then map them to our 33 software markets to approximate the overall sales growth for each market for a given year.

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 and measured it by the claims-weighted patent count related to each software market j cumulated up to year t.

Patent quality. This variable is a control for the quality of patents in the market j in year t. As has been noted elsewhere, higher quality patents suggest superior technological capabilities possessed by existing incumbents in the market, 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 market j (adjusted for truncations) divided by total number patents in j up to year t.

Open Invention Network (OIN) patents. At the end of our sample period, another institution similar to patent commons—OIN—was established. Similar to The Commons, OIN offers contractually royalty-free usage of its patents to OSS participants as long as users promise not to file suit against software associated with the Linux System.[26] We do not focus on this institution in our main analysis as it was introduced too late in our sample period to have a measureable effect.[27] However, we include it as a control. We measure this variable as the claims-weighted patent count of OIN patents related to software market 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

[25] 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.
[26] For the detailed definition of the Linux system, see http://www.openinventionnetwork.com/pat_linuxdef.php.
[27] For the 130 patents contributed to OIN from year 2006 to 2009, 70 percent were contributed in 2008 and 2009.

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 on 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)[28]. We compute the claims-weighted patent count of the SSO patents that are distributed under royalty-free licenses and are related to software market j cumulated up to year t.

OSS demand. While *Sales growth* could potentially control for the market growth for software market j in year t in general, the market demand for OSS products may deviate from the demand for proprietary software products. We capture the demand for OSS using the total OSS downloads at SourceForge.net (SourceForge), which is the largest repository for OSS projects over the world. We first match over 0.2 million OSS projects at SourceForge to the 33 software markets according to the projects' categories. Then we measure this variable by computing the total number of downloads for all OSS applications related to software market j cumulated up to year t.

6. Empirical Strategies and Results

We motivate our empirical analyses by first investigating the distribution of patents in The Commons across various software markets and the value of patents in The Commons. Next, we establish a baseline relationship between OSS entry and the size of The Commons, and demonstrate how the marginal impact of The Commons varies with the cumulativeness of innovation or patent ownership concentration. We then show that our baseline results are robust to a variety of specifications and robustness checks.

6.1. The distribution of The Commons' patents across software markets and their quality

Figure 1 shows the distribution of patents in The Commons across 33 software markets. Operating systems and utilities market, disk/file management market, and database software market have the greatest concentration, followed by software development tools and system management software market. Because roughly 95% of patents in The Commons were contributed by IBM, we compare this to the distribution of IBM's patents across markets. The two distributions are quite similar. In the Appendix

[28] 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.

Figure C-1, we show that the fraction of IBM patents that are contributed to The Commons are quite similar across markets.

We next examine the quality of patents in The Commons relative to comparison groups. We first compare patents in The Commons to similar market patents. Following the matching method employed by Thompson and Fox-Kean (2005), we pair each patent in The Commons with a randomly selected control patent from the same market that matches the primary classification of The Commons' patent at the subclass level and that was applied for in the same year or within a one year window. As shown in Table 2a, there is no statistically significant different between the two groups in terms of forward citations or backward citations. However, patents in the Commons patent seem to cover a narrower technology scope, as measured by the number of claims.

We next compare the quality of IBM's donated patents to other patents in its portfolio using the same procedure described above. As shown in Table 2b, there is no significant difference in forward citations between the two groups. However, patents that are not in The Commons seem to have a significantly higher number of backward citations and claims. This suggests that while IBM may be contributing patents covering a narrower technology space, the patents in The Commons are less derivative than IBM's other comparable patents.

6.2. Baseline results and robustness checks

We first seek to establish whether increases in donated IPRs are associated with increasing entry in related markets. Our estimation framework is motivated by recent research that has studied how patent thickets influence market entry in the software industry (e.g., Cockburn and MacGarvie 2011). We model new product entry using count data models with conditional fixed effects. Suppose the number of OSS entrants in software market *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 market-specific and time-constant variable that incorporates unobserved heterogeneity across markets. Thus, $E(Y_{jt} \mid X_{jt}, \alpha_j) = \lambda_{jt} = \alpha_j \exp(X_{jt}'\beta)$, and to test H1 we assume

$$X_{jt}'\beta = \beta_1 \textit{The Commons}_{jt} + \gamma_1 \textit{PatentThicket}_{jt-1} + \gamma_2 \textit{SalesGrowth}_{jt} + \gamma_3 \textit{MarketPatents}_{jt-1} + \gamma_4 \textit{OtherFreePatents}_{jt} + \gamma_5 \textit{OSSdemand}_{jt} + \tau_t \quad (1)$$

The vector $\textit{PatentThicket}_{jt-1}$ includes the two patent thicket variables $\textit{cumulativeness}_{jt-1}$ and $\textit{concentration}_{jt-1}$; the vector $\textit{MarketPatents}_{jt-1}$ includes the $\textit{Total patents}_{jt-1}$ and $\textit{Patent quality}_{jt-1}$.The two vectors are lagged by one year to allow for any lagged effects on OSS entry. The vector $\textit{OtherFreePatents}_{jt}$ represents the patents from OIN and SSOs—$\textit{OIN patents}_{jt}$ and $\textit{SSO patents}_{jt}$. 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 market level. We are interested in the estimate for β_1 which, if positive, supports hypothesis 1. Our identification assumption in this section is that there are no unobserved market-level factors that are correlated both with new market entry and with the number of patents in The Commons. This assumption will be violated if, for example, IBM and other contributors to The Commons donate the greatest number of patents in markets that are growing rapidly (and this growth is not included in our control variables). We will examine the implications of weakening this assumption in a variety of ways in the analyses below.

Columns (1) and (2) in Table 3 report the estimation results for specification (1). Column (1) shows the specification with baseline controls including the $SalesGrowth_{jt}$, $PatentThicket_{jt-1}$, and $MarketPatents_{jt-1}$ as well as with market and year fixed effects. We include sales growth in all specifications because both demand and market competition within a market are important determinants of start-up entry.[29] Meanwhile, because of the central role of the vector $PatentThicket_{jt-1}$ in our hypotheses and the highlighted effect from $MarketPatents_{jt-1}$ on start-up entry by the literature (e.g., Cockburn and MacGarvie 2011), the two patent-related vectors are incorporated in all specifications as well. The coefficient in column (1) suggests that a 10% increase in The Commons' patent claims related to a software market is associated with a 1.3% increase in OSS entry in that market. Results are robust when we add other controls such as $OtherFreePatents_{jt}$ and $OSSdemand_{jt}$, as shown in column (2) in Table 3. We note that while increases in the size of The Commons are associated with OSS entry, increases in the size of royalty-free patents in SSOs are not. We speculate that this may reflect differences in the licensing requirements for The Commons and SSO patents: in particular, while users of The Commons pledge not to sue The Commons' 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.

We next investigate whether the impact of The Commons is greater in markets with high cumulativeness of innovation and the concentration of patent ownership in a market. Our key identification assumption here is somewhat weaker—that growth in claims-weighted patents in The Commons are not correlated with unobservable factors that influence entry that are differentially trending in markets with high cumulativeness and concentration. We begin by examining whether the impact of The Commons is higher when the cumulativeness of innovation in a market is high. The specification for $X_{jt}'\beta$ becomes

$$X_{jt}'\beta = \beta_1 \mathit{The\ Commons}_{jt} + \beta_2 \mathit{The\ Commons}_{jt} * \mathit{cumulativeness}_{jt-1} + \gamma_1 \mathit{PatentThicket}_{jt-1} + \gamma_2 \mathit{SalesGrowth}_{jt} + \gamma_3 \mathit{MarketPatents}_{jt-1} + \gamma_4 \mathit{OtherFreePatents}_{jt} + \gamma_5 \mathit{OSSdemand}_{jt} + \tau_t. \quad (2)$$

[29] 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.

As shown in columns (3) and (4) in Table 3, a 10% increase in the size of The Commons is associated with a 2.8%-3% increase in OSS entry, with the effect computed at the average level of cumulativeness of innovation. Further, while the marginal effect of The Commons is insignificant when evaluated at the 10^{th} percentile of the cumulativeness, the effects are statistically and economically significant when evaluated at the 90^{th} percentile. Specifically, a 10% increase in the size of The Commons is associated with a 5.2%-5.4% increase in OSS entry when cumulativeness of innovation is at its 90^{th} percentile. A test for the difference of the two marginal effects (at the 10^{th} and 90^{th} percentiles) is statistically significant. As described earlier, we also constructed alternative measures of the cumulativeness of innovation and our results are robust to these changes.[30] The results suggest that the impact of The Commons is greater when the cumulativeness of innovation is high.

Similarly, to explore how the impact of The Commons is influenced by the variation in concentration of patent ownership, the specification for $X_{jt}'\beta$ can be written as

$$X_{jt}'\beta = \beta_1 \textit{The Commons}_{jt} + \beta_2 \textit{The Commons}_{jt} * \textit{concentration}_{jt-1} + \gamma_1 \textit{PatentThicket}_{jt-1} + \gamma_2 \textit{SalesGrowth}_{jt} + \gamma_3 \textit{MarketPatents}_{jt-1} + \gamma_4 \textit{OtherFreePatents}_{jt} + \gamma_5 \textit{OSSdemand}_{jt} + \tau_t. \quad (3)$$

The empirical results for this specification are reported in columns (5) and (6) in Table 3. While the marginal effect of The Commons is insignificant when concentration is at its 10^{th} percentile or mean value, a 10% increase in the size of The Commons is associated with a 1.1%-1.2% increase in OSS entry when concentration is at its 90^{th} percentile. The test for the difference of marginal effects of The Commons between concentration evaluated at the 10^{th} percentile and the 90^{th} percentile is statistically significant at the 10% level.

To present a more complete picture of how the impact of The Commons 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) and (8) in Table 3. A 10% increase in the size of The Commons is associated with a 4.4%-4.5% increase in OSS entry when the cumulativeness of innovation is at its 90^{th} percentile; the marginal effect of The Commons is significantly different at the 1% level when evaluated at high versus low level of cumulativeness. In this specification, while there is no statistically significant difference between the marginal effects evaluated at the 10^{th} and 90^{th} percentiles of concentration, the sign for the interaction between The Commons and patent ownership concentration remains positive across specifications. We suspect this non-significance is likely caused by the multicollinearity between the two interactions.[31] Nevertheless, it does suggest while the qualitative nature of our results is similar when including cumulativeness and concentration together, the statistical significance of the concentration result is weaker.

[30] The regression results based on the robust measure of cumulativeness are available upon request.
[31] In the pooled sample, the simple correlation coefficient between the two interaction terms is 0.66.

We implemented a series of additional analyses to study the robustness of our results. We estimated all of our models using OLS models with market fixed effects. We substituted our claims-weighted count of number of Commons patents in the market with a raw patent count. Because of the decline in the rate of new OSS product entry after 2006 (in Figure 5), we re-estimated the baseline model using a sample endpoint of 2007 and 2008. As detailed in Appendix C Tables C-1 through C-4, our results are robust to all of these changes.

Because our measure of patent ownership concentration could be confounded with the concentration of market structure, we constructed a measure of market structure concentration using the share of top 4 incumbents' sales in each market in each year. We found that adding this measure as an additional control did not affect our finding that The Commons had a stronger positive effect on entry in markets with high patent concentration. Further, we interacted the new market structure control with The Commons variable, in regressions both including and excluding our interaction of concentration of patent ownership with The Commons. In both Poisson regression (in Table C-5) and linear models (in Table C-6) we find that changes in market structure concentration have no significant impact on the marginal effect of increases in the size of The Commons. However, even when controlling for the effects of market structure, changes in The Commons have a stronger effect on entry when patent concentration is high. This set of tests supports the view that concentration of IPR holdings, not concentration of market structure, influence the marginal effect of The Commons on OSS entry.

6.3. Omitted variable bias

This section provides the results of a variety of additional tests we run to address omitted variable bias and simultaneity. We first employ a model that includes a linear time trend interacted with the market fixed effects. The results are presented in Table 4. In column 1 we show the results of a model with market-specific time trends but excluding the interactions with cumulativeness and concentration. The results are qualitatively similar to our baseline, though the size of the marginal effect is somewhat smaller. We next interact the size of The Commons with cumulativeness and concentration. Consistent with the baseline results, while the marginal effect of The Commons is insignificant when either cumulativeness of innovation or concentration is at its 10th percentile, a 10% increase in the size of The Commons is associated with a 2%-4% increase in OSS entry when cumulativeness is at its 90th percentile and is associated with a 1.1%-1.5% increase in OSS entry when concentration is at its 90th percentile, as shown by columns (2) through (4). The test of the difference between high and low cumulativeness and the test of the difference between high and low concentration are also largely consistent with the baseline results.

We next discuss the results of a series of instrumental variable estimates. We employ a count data model with instrumental variables estimated using Generalized Method of Moments (GMM) estimation. To test whether The Commons has positive effect on OSS entry, 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):

$$Y_{jt} = \alpha_j\exp(X_{jt}'\beta) + u_{jt} = \alpha_j \exp(\beta_1 \textit{The Commons}_{jt} + \gamma_1 \textit{PatentThicket}_{jt-1} + \gamma_2 \textit{SalesGrowth}_{jt} + \gamma_3 \textit{MarketPatents}_{jt-1} + \gamma_4 \textit{OtherFreePatents}_{jt} + \gamma_5 \textit{OSSdemand}_{jt} + \tau_t) + u_{jt} = \mu_{jt}\alpha_j + u_{jt} \qquad (4)$$

where $\mu_{jt} = \exp(\beta_1 \textit{The Commons}_{jt} + \gamma_1 \textit{PatentThicket}_{jt-1} + \gamma_2 \textit{SalesGrowth}_{jt} + \gamma_3 \textit{MarketPatents}_{jt-1} + \gamma_4 \textit{OtherFreePatents}_{jt} + \gamma_5 \textit{OSSdemand}_{jt} + \tau_t)$ and u_{jt} is the error term. *The Commons*$_{jt}$ is treated as a potentially endogenous variable. Suppose $X_{jt}\gamma = \gamma_1 \textit{PatentThicket}_{jt-1} + \gamma_2 \textit{SalesGrowth}_{jt} + \gamma_3 \textit{MarketPatents}_{jt-1} + \gamma_4 \textit{OtherFreePatents}_{jt} + \gamma_5 \textit{OSSdemand}_{jt} + \tau_t$, 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 market-specific fixed effect, and obtain the following moment condition:

$$E\left(Z_{jt}\left(\frac{Y_{jt}}{\mu_{jt}} - \frac{Y_{jt-1}}{\mu_{jt-1}}\right)\right) = 0 \qquad (5)$$

where Z_{jt} includes the set of exogenous variable X_{jt} and a set of instruments for *The Commons*$_{jt}$ 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 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 $\bar{z} = \frac{1}{NT}\sum_{j=1}^{N}\sum_{t=1}^{T} z_{jt}$.

Since 95% of the patents in The Commons were contributed by IBM in January 2005, our choice of instrumental variables (IVs) mostly focuses on factors that influence IBM's decision on what to contribute and the timing of the contribution behavior. Our first instrumental variable (IV) is the cumulated number of IBM patents litigated at the European Patent Office (EPO) up to year *t* and related to software market *j* (denoted as *IBM litigated patents*). This IV is closely related to IBM's holdings of blocking patents across different markets, which could influence IBM's decision on what to contribute to The Commons for the following reason. As noted earlier, the existence of blocking patents could increase the sunk costs of entry by start-ups, so having its blocking patents pledged would be the most effective way for the incumbent to encourage entry in order to benefit from complementary products or services. While we could use all IBM's litigated patents to proxy for its propensity to create blocking effects, we use IBM's patents litigated at the EPO in particular, as the litigation event filed at the EPO should be less correlated with new OSS product entries in the US market. In order to be consistent with how we measured The Commons, we use claims-weighted patent count.

We then interact IBM litigated patents with a time dummy that turns on after year 2003 (denoted as *afteryear2003*) to construct our second IV (denoted as *IBM litigated patents X afteryear2003*). The motivation to use *afteryear2003* 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 firms developing OSS-related product and attracted significant publicity. As such, it is expected to influence the timing of IBM's patent contribution decision. Moreover, because this litigation event against IBM is in particular related to its deployment of Linux operating system, we further interact this second IV with a dummy that turns on for operating system market to form our third IV (denoted as *IBM litigated patents X afteryear2003 X Linux-related market*).

Table 5 shows the results of an auxiliary first stage OLS regression of the number of claims-weighted patents on each of our instruments separately and then together. The direction of the coefficient estimates is as expected; the number of claims-weighted Commons patents is increasing in the number of litigated IBM patents, and this is particularly so after 2003 and in markets related to the Linux operating system. The first stage F-statistics range between 8.94 and 53.47, suggesting the IVs have some power in explaining the endogenous variable.

Table 6 reports the GMM estimation results where we instrument for *The Commons* using all three instruments. As shown in column (1) in Table 5 the coefficient is significantly positive, consistent with our baseline estimates[32]. The result from Hansen J statistic fails to reject the null that the instruments used are uncorrelated with the error term. The remaining columns in Table 6 show the results of instrumental variables regressions in which the endogenous variables are *The Commons* interacted with *Cumulativeness* and *Concentration* and the instruments are the three instruments described above and their interaction with *Cumulativeness* and *Concentration.* The estimated coefficients for the two interaction variables remain significantly positive across all specifications. For robustness, we also run GMM estimation using each of the IVs. The results are presented in Table C-7 in the Appendix and are qualitatively similar to the estimation results based on the full set of IVs. While the combined empirical evidence from Tables 3 through 5 largely confirms the greater effect of The Commons in markets with highly concentrated patent ownership, it seems to suggest that the interaction between The Commons and concentration of patent ownership is not as strong as its interaction with cumulativeness.

6.4. Falsification exercise

As suggested in the earlier graphs 2 through 4, markets with few patents in The Commons show little variation in year to year entry of OSS products. In contrast, in markets with a large number of Commons patents, there is a significant increase in entry in the years immediately following the introduction of The Commons in 2005. Figure 5 complements Figure 2 and shows the time series of new

[32] In this set of GMM estimations, we only use baseline controls (i.e. the controls used in column (1) in Table 3), as adding more controls tends to lead to non-convergent results.

entry for markets with small and large size of The Commons. It suggests a significant increase in entry in 2004, one year prior to the introduction of The Commons. We expect that this result reflects the combination of other events occurring in the OSS community during this time period, as well as other actions taken by IBM in support of OSS. In particular, as mentioned earlier, one significant event in the OSS community during this time period was the March 2003 lawsuit filed against IBM for alleged copyright infringement in IBM's deployment of the Linux code case. This lawsuit, together with a well-publicized report in 2004 that suggested that Linux was potentially infringing on several hundred patents[33], initiated a series of actions which eventually lead to the introduction of The Commons and subsequent actions taken to provide additional legal protections for OSS firms against patent infringement lawsuits.[34] For IBM, this included the initiation of a legal defense fund in Linux in January 2004 and IBM Senior Vice President Nick Donofrio's announcement that IBM would not assert its patents against the Linux kernel in August 2004.

To investigate how these other events may have affected OSS entry, we construct a variable, denoted as "*False Commons*", by interacting the patent density in The Commons across markets with a time dummy that turns on for years 2003 and 2004. If we add this *False Commons* to the specification, as shown in column (1) in Table 7, in line with the time-series chart in Figure 5, the coefficient for *False Commons* is significantly positive. To capture how other non-Commons activities by IBM during this period might affect OSS entry, we add to regression model (1) the interaction of the cumulative number of claims-weighted *IBM patents* in the market that are not in The Commons and its interaction with a dummy that turns on for years 2003 and 2004 (denoted as *IBM patents* X *2003_2004*) As shown in column (2), both *IBM patents* and *IBM patents* X *2003_2004* are significantly and positively correlated with OSS entry, reflecting how IBM's other actions during this time period affected OSS entry. More importantly, as shown in column (3), after controlling for the effect of IBM's IPR commitments and its actions in 2003 and 2004, *False Commons* becomes insignificant while *The Commons* (whose effect "turns on" in 2005) remains significantly positive. These findings corroborate our hypothesize effect, suggesting that IBM's pledge to The Commons go beyond these earlier OSS support actions and announcements, by identifying a specific set of complementary technologies that may be valuable to Linux and existing open source projects in which IBM will not assert its patents.

We implement an additional falsification exercise that examines the relationship between The Commons and proprietary software product entry. Given our theoretical predictions, we should not observe a positive effect of The Commons on entry of proprietary software products. Focusing on PROMT articles assigned with PROMT product codes, we identified entry with proprietary product

[33] See Alexy and Reitzig (2013) for detailed discussion about this event.
[34] Including, but not limited to, the introduction of the Open Invention Network in November, 2005.

introductions by start-ups into 29 software markets from year 2002 to year 2009.[35] The results, as reported in Table 8, suggest that The Commons has no significant effect on proprietary entry; further, the interaction of The Commons with cumulativeness or concentration plays little role in proprietary entry as well. However, these results need to be interpreted with caution, as the proprietary entry events were identified through PROMPT articles assigned with product codes so we might miss product introduction events contained in articles without product codes.

7. Conclusions

In this study we provide the first large sample evidence of how the provision of patent commons shapes OSS entry by start-ups. In particular, we examine the introduction of The Commons, a patent commons initiated by the Open Source Development Labs and IBM in 2005. We show that increases in the size of The Commons related to a software market are associated with increases in OSS entry by start-up software firms in that market. Furthermore, the impact of The Commons is magnified when two features of patent thickets are present in the market: cumulativeness of innovation and concentration of patent ownership. We observe a particularly strong relationship between the size of The Commons and OSS start-ups' entry in markets with high cumulativeness of innovation; the marginal effect of The Commons is also greater when concentration is high, although this result is not robust across all specifications.

Our study is subject to limitations. In particular, our study analyzes the impact of patent commons 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.[36] The introduction of patent commons 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 patent commons 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.

Our study has implications for our understanding of the impact of formal IPR on OSS, an area in which our knowledge is still quite limited (Lerner and Tirole 2005a, von Hippel and von Krogh 2003). In particular, our research adds to the literature that looks at how IPR licensing and enforcement influences OSS innovation (Graham and Mowery 2005, Lerner and Tirole 2005b, Maurer and Scotchmer 2006) as

[35] Our use of the aggregated 29 markets (rather than the baseline 33 markets) from 2002 to 2009 reflects our data constraints. More details are provided in Table 8.

[36] A natural and interesting extension to this study is to look at how the size of The Commons 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 Commons on the survival of start-up firms who produce OSS.

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).

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 patent commons 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).

References

Alexy, O., M. G Reitzig. 2013. Private–collective Innovation, Competition, and Firms' Counterintuitive Appropriation Strategies. *Research Policy* 42 (4): 895-913.

Al-Najjar, N., R. Smorodinsky. 2000. Pivotal Players and Characterization of Influence. *Journal of Economic Theory* 92: 318-342.

Arora, A., A. Fosfuri, and A. Gambardella. 2001. *Markets for Technology: Economics of Innovation and Corporate Strategy.* Cambridge: MIT Press.

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

Auza, Jun. 2011. Infamous Microsoft FUD Campaigns Against Linux. Available at www.junauza.com, posted on July 18, 2011.

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

Bessen, J. and M. J. Meurer. 2008. *Patent Failure: How Judges, Bureaucrats, and Lawyers Put Innovators at Risk.* Princeton: Princeton University Press.

Biddle, B., A. White, and S. Woods. 2012. How Many Standards in a Laptop? (And Other Empirical Questions) Working Paper, Sandra Day O'Connor College of Law, Arizona State University.

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.

Campbell-Kelly, M., D.D. Garcia-Swartz. 2009. Pragmatism, not ideology-Historical perspectives on IBM's adoption of open-source software. *Information Economics and Policy*. 21(3): 229–244.

Capek, P. G., S. P. Frank, S. Gerdt, and D. Shields. 2005. A history of IBM's open-source involvement and strategy. IBM Systems Journal 44(2): 249-257.

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

Coase, R. 1960. The Problem of Social Cost. *Journal of Law and Economics* 3(1): 1-44.

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.

Gans, J. D. Hsu, and S. Stern. 2002. When does start-up innovation spur the gale of creative destruction? *RAND Journal of Economics* 33(4): 571-586.

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.

Hagiu, A. and D. Yoffie. 2013. The New Patent Intermediaries: Platforms, Defensive Aggregators, and Super-Aggregators. *Journal of Economic Perspectives* 27(1); 45-66.

Hall, B.H., Helmers, C. Forthcoming. Innovation and Diffusion of Clean/Green Technology: Can Patent Commons Help? *Journal of Environmental Economics and Management*.

Hall, B.H., C. Helmers, G. von Graevenitz, and C. Rosazza-Bondibene. 2012. A Study of Patent Thickets: Final report prepared for the UK Intellectual Property Office.

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.

Hall, B. H. and R. H. Ziedonis. 2001. The patent paradox revisited: An empirical study of patenting in the U.S. semiconductor industry, 1979-1995. *RAND Journal of Economics* 32(1): 101-128.

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.B. and M. Trajtenberg. 2004. *Innovation and its Discontents: How Our Broken Patent System is Endangering Innovation and Progress, and What to Do About it.* Princeton: Princeton University Press.

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 Commons 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 Commons 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 Commons 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 Commons: The Determinants of Licensing Rules. *The RAND Journal of Economics* 38(3): 610-25.

Lerner, J., J. Tirole. 2004. Efficient patent Commons. *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.

Mailath, G. and A. Postlewaite. 1990. Asymmetric Information Bargaining Problems with Many Agents. *Review of Economic Studies* 57(3): 351-367.

Marson, Ingrid. 2004. Linus Torvalds speaks out against EU Patent Law. ZDNet. Available at http://www.zdnet.com/linus-torvalds-speaks-out-against-eu-patent-law-3039174746/. Accessed April 30, 2013.

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.

Musil, Steven. 2011. Apple, RIM in group buying Nortel patents for $4.5B. Cnet news. Available at http://news.cnet.com/8301-1001_3-20075977-92/apple-rim-in-group-buying-nortel-patents-for-$4.5b/. Accessed April 30, 2013.

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.

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.

Samuleson, P. 2006. IBM's pragmatic embrace of open source. *Communications of the ACM*, 49(10): 21-25.

Scannell, Ed. 2004. IBM says it won't assert patents against Linux kernel. InfoWorld Daily News, August 4, 2004.

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

Schultz, Jason and Jennifer M. Urban. 2012. Protecting Open Innovation: The Defensive Patent License as a New Approach to Patent Threats, Transaction Costs, and Tactical Disarmament. *Harvard Journal of Law and Technology* 26(1): 1-67.

Scotchmer. 2004. *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 Commons Demonstrates Variety of Purposes and Management Structures. KEI Research Note 6.

Shapiro, C. 2001. Navigating the Patent Thicket: Cross Licenses, Patent Commons, 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.

Simcoe, T. 2006. Open Standards and Intellectual Property Rights, H. Chesbrough, W. Vanhaverbeke, and J. West, eds. *Open Innovation, Research a New Paradigm*. Oxford: Oxford University Press.

Stallman, R. 2011. Europe's "unitary patent" could mean unlimited software patents. Available at http://www.gnu.org/philosophy/europes-unitary-patent.html. Access April 30, 2013.

Teece, D. 1986. Profiting from technological innovation: Implications for integration, collaboration, licensing, and public policy. *Research Policy* 15(6): 285-305.

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.

Table 1: Summary Statistics

Variable name	Measure (Market-year)	Obs.	Mean	Std. Dev.	Min	Max
Dependent variable						
OSS entry	The number of new OSS entrants into market j in year t	363	.667	1.344	0	11
Independent variables and controls						
The Commons	Log of The Commons' claims-weighted patent count related to market j cumulated up to year t	363	2.413	2.936	0	7.911
Cumulativeness	Log of cumulativeness of innovation in market j up to year t	363	.808	.613	.095	3.454
Concentration	Four-assignee citation concentration ratio in market j up to year t	363	0.227	0.075	0.076	0.458
Sales growth	The growth of market j's sales from year t-1 to year t	363	1.007	0.150	0.631	2.112
Total patents	Log of total claims-weighted patent count related to market j cumulated up to year t	363	10.817	1.232	6.870	13.486
Patent quality	Log of average quality of patents related to market j cumulated up to year t	363	2.832	.402	1.839	4.051
OIN patents	Log of Open Invention Network's claims-weighted patent count in market 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 market j cumulated up to year t	363	1.628	2.118	0	5.908
OSS demand	Log of cumulated downloads at SourceForge for OSS applications related to market j up to year t	363	14.340	5.721	0	22.365
Instrument variables						
IBM litigated patents	IBM's patents litigated at the European Patent Office (EPO) related to market j cumulated up to year t	363	91.463	87.480	0	401
IBM litigated patents X afteryear2003	IBM's patents litigated at the European Patent Office (EPO) related to market j cumulated up to year t * afteryear2003 dummy, where afteryear2003 dummy turns on for year t = 2003, 2004, …, 2009	363	65.331	90.787	0	401
IBM litigated patents X afteryear2003 X Linux-related market	IBM's patents litigated at the European Patent Office (EPO) related to market j cumulated up to year t * after_year2003 dummy * Linux-related market, where Linux-related market dummy turns on for "operating systems and utilities" market	363	5.132	37.393	0	351

Table 2a: Patents in The Commons compared to other similar market patents

		Patents in The Commons	Other similar market patents	T-test
	Obs.	517	517	
Forward citations as of Dec 2004	Mean (Std.Err.)	11.393 (.683)	12.334 (.788)	-0.903
Backward citations	Mean (Std.Err.)	9.718 (.314)	9.739 (.394)	-0.042
Claims	Mean (Std.Err.)	17.321 (.496)	19.400 (.686)	-2.458***

Note: 1) For each of the patents in The Commons, we paired it by randomly selecting a control patent from all market patents that matched the primary classification of The Commons patent at the subclass level and that were applied in the same year or within one-year window of The Commons patent. 2) To eliminate any effect of contributing to The Commons on the increase in forward citation, when comparing the forward citations, we only count the total forward citations received by December 2004, as The Commons was formed in January 2005. Because 57% of the patents in Commons are granted before 1998 and 97% of the patents are granted before 2002, we believe the truncation on forward citations is not a major issue in this setting. 3) ***: significant at 1%.

Table 2b: IBM's Patents in The Commons compared to IBM's other similar patents

		IBM's Patents in The Commons	IBM's other similar patents	T-test
	Obs.	407	407	
Forward citations as of Dec 2004	Mean (Std.Err.)	11.044 (.784)	10.341 (.702)	0.668
Backward citations	Mean (Std.Err.)	9.683 (.358)	11.545 (.509)	-2.992***
Claims	Mean (Std.Err.)	16.813 (.501)	18.646 (.612)	-2.319**

Note: 1) For each of IBM's patents in The Commons, we paired it by randomly selecting a control patent from all IBM's patents that matched the primary classification of the focal patent at the subclass level and that were applied within one-year window of The Commons patent. 2) To eliminate any effect of contributing to The Commons on the increase in forward citation, when comparing the forward citations, we only count the total forward citations received by December 2004, as The Commons was formed in January 2005. Because all IBM's patents in Commons are granted in year 1993, 1997, or 2001, we believe the truncation on forward citations is not a major issue in this setting. 3) ***: significant at 1%, **: significant at 5%.

Table 3: Baseline results, conditional fixed-effect Poisson regression

Dependent variable: OSS entry	Specification testing direct impact of The Commons only		Add interaction with cumulativeness of innovation		Add interaction with concentration of patent ownership		Add interaction with cumulativeness of innovation and with concentration of patent ownership	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
The Commons	.127* (.083)	.145* (.087)	.033 (.119)	.052 (.124)	-.079 (.114)	-.098 (.135)	-.080 (.120)	-.071 (.164)
The Commons * Cumulativeness			.330*** (.096)	.331*** (.098)			.308*** (.095)	.306*** (.100)
The Commons * Concentration					.566* (.322)	.660* (.369)	.309 (.270)	.336 (.377)
Cumulativeness	.965 (.977)	1.010 (1.156)	3.194*** (.858)	3.303*** (.983)	1.063 (.886)	1.030 (1.024)	3.116*** (.854)	3.142*** (.990)
Concentration	-6.961 (5.273)	-7.625 (7.219)	-11.238** (4.812)	-12.196* (7.097)	-11.293* (6.552)	-10.709 (7.459)	-13.251** (5.539)	-13.375* (6.878)
Sales growth	-.412 (.338)	-.400 (.366)	-.571* (.379)	-.567 (.406)	-.454 (.354)	-.470 (.381)	-.585* (.384)	-.590 (.415)
Total patents	-.707 (1.408)	-.713 (1.505)	.797 (1.430)	.852 (1.465)	-1.096 (1.443)	-1.157 (1.560)	.510 (1.439)	.527 (1.519)
Patent quality	-1.435 (2.453)	-1.599 (2.387)	-1.224 (2.401)	-1.361 (2.335)	-1.529 (2.523)	-1.684 (2.422)	-1.284 (2.445)	-1.401 (2.350)
OIN patents		-.066 (.071)		-.069 (.065)		-.076 (.070)		-.073 (.066)
SSO patents		-.010 (.078)		-.016 (.065)		.029 (.083)		.005 (.078)
OSS demand		.062 (.095)		.055 (.083)		.062 (.100)		.056 (.086)
Year dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Market fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	286	286	286	286	286	286	286	286
Log pseudolikelihood	-227.909	-227.351	-223.884	-223.327	-226.961	-226.270	-223.630	-223.082
Marginal Effects								
The Commons (average)	.127* (.083)	.145* (.087)	.278** (.115)	.297** (.120)	.046 (.072)	.048 (.081)	.217** (.085)	.230** (.114)
The Commons (cumulativeness=10%)			.095 (.114)	.114 (.119)			.047 (.091)	.061 (.110)
The Commons (cumulativeness=90%)			.519*** (.149)	.540*** (.154)			.442*** (.119)	.454*** (.153)
Test of the difference between high and low cumulativeness, p-value			.001	.001			.001	.002
The Commons (concentration =10%)					-.005 (.084)	-.011 (.098)	.190** (.091)	.200* (.134)
The Commons (concentration =90%)					.104* (.071)	.115* (.074)	.249*** (.086)	.265*** (.100)
Test of the difference between high and low concentration, p-value					.078	.073	.252	.372

Notes: 1) Robust standard errors, clustered by market, are in parentheses. 2) * significant at 10%, ** significant at 5%, *** significant at 1%. 3) The number of observations is lower than 363 because of the use of conditional fixed effects Poisson models, which drops markets without OSS entry over the entire sample period.

Table 4: Robustness test using market-specific time trends, conditional fixed-effect Poisson regression

Dependent variable: OSS entry	Specification testing direct impact of The Commons only	Add interaction with cumulativeness of innovation	Add interaction with concentration of patent ownership	Add interaction with cumulativeness of innovation and with concentration of patent ownership
	(1)	(2)	(3)	(4)
The Commons	.093** (.041)	-.048 (.079)	-.235** (.102)	-.267 (.183)
The Commons * Cumulativeness		.298** (.118)		.172* (.112)
The Commons * Concentration			1.072*** (.389)	.909 (.663)
Cumulativeness		.612 (1.380)		-.273 (1.323)
Concentration			17.016** (8.210)	17.734 (7.912)
Market-specific time trend	Yes	Yes	Yes	Yes
Market fixed effects	Yes	Yes	Yes	Yes
Observations	286	286	286	286
Log pseudolikelihood	-231.251	-229.011	-223.827	-223.018
Marginal Effects				
The Commons (average)	.093** (.041)	.174*** (.045)	.002 (.044)	.062 (.057)
The Commons (cumulativeness=10%)		.009 (.061)		-.033 (.067)
The Commons (cumulativeness=90%)		.391*** (.114)		.188* (.115)
Test of the difference between high and low cumulativeness, p-value		.012		.124
The Commons (concentration =10%)			-.093 (.060)	-.019 (.105)
The Commons (concentration =90%)			.112** (.055)	.155*** (.052)
Test of the difference between high and low concentration, p-value			.005	.170

Notes: 1) Robust standard errors, clustered by market, are in parentheses. 2) * significant at 10%, ** significant at 5%, *** significant at 1%. 3) "Market-specific time trend" is measured by a linear time trend times 33 market dummies. 4) Given our sample size with only 286 observations, it is difficult to incorporate any other controls besides these 33 market-specific time trends.

Table 5: Results of OLS regression of Claims-weighted patents in The Commons on instruments

Dependent variable: The Commons	IV: IBM litigated patents	IV: IBM litigated patents X afteryear2003	IV: IBM litigated patents X afteryear2003 X Linux-related market	The full set of IVs
	(1)	(2)	(3)	(4)
IBM litigated patents	.008*** (.001)			-.001 (.003)
IBM litigated patents X afteryear2003		.005*** (.001)		.005*** (.001)
IBM litigated patents X afteryear2003 X Linux-related market			.003*** (.001)	.001 (.001)
First Stage F-statistic, p-value	27.89 (0.00)	53.47 (0.00)	8.94 (0.00)	18.99 (0.00)
Controls	*SalesGrowth*, *MarketPatents*, *PatentThicket*	*SalesGrowth*, *MarketPatents*, *PatentThicket*	*SalesGrowth*, *MarketPatents*, *PatentThicket*	*SalesGrowth*, *MarketPatents*, *PatentThicket*
Year dummies	Yes	Yes	Yes	Yes
Market fixed effects	Yes	Yes	Yes	Yes

Notes: 1) The full set of IVs include: IBM litigated patents, IBM litigated patents X afteryear2003, and IBM litigated patents X afteryear2003 X Linux-related market. 2) 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, are in parentheses. 6) * significant at 10%, ** significant at 5%, *** significant at 1%.

Table 6: GMM estimates of count data regression using the full set of IVs

Dependent variable: OSS entry	Specification testing direct impact of The Commons only	Add interaction with cumulativeness of innovation	Add interaction with concentration of patent ownership	Add interaction with cumulativeness of innovation and with concentration of patent ownership
	(1)	(2)	(3)	(4)
The Commons	.240* (.157)	-.007 (.111)	-.311 (.210)	-.853*** (.168)
The Commons * Cumulativeness		.303*** (.100)		.519*** (.139)
The Commons * Concentration			1.790** (.773)	2.577*** (.807)
Cumulativeness	-2.513 (2.752)	-2.423* (1.539)	.238 (.834)	3.628*** (.696)
Concentration	3.791 (6.606)	-.728 (4.276)	-5.097 (4.720)	-8.492* (5.310)
Sales growth	2.331** (1.080)	1.243*** (.479)	.066 (.463)	-.342 (.444)
Total patents	-.024 (1.276)	.121 (.714)	.941** (.398)	2.932*** (.491)
Patent quality	2.488 (2.085)	4.330*** (.794)	1.11 (.968)	1.283 (1.082)
Year dummies	Yes	Yes	Yes	Yes
Observations	363	363	363	363
Over-identification, p-value	0.22	0.17	0.10	0.19

Notes: 1) The full set of IVs include: IBM litigated patents, IBM litigated patents X afteryear2003, and IBM litigated patents X afteryear2003 X Linux-related market. 2) Robust standard errors, clustered by market, are in parentheses. 3) * significant at 10%, ** significant at 5%, *** significant at 1%. 4) We use Wooldridge's quasi-differencing transformation to remove market fixed effects.

Table 7: Falsification test by adding "False Commons", conditional fixed-effect Poisson regression

Dependent variable: OSS entry	(1)	(2)	(3)
The Commons	.373*** (.117)	.383*** (.140)	.426*** (.134)
False Commons	.463*** (.137)		.260 (.200)
IBM patents		.908*** (.249)	.909*** (.257)
IBM patents X 2003_2004		.565*** (.160)	.308 (.251)
Cumulativeness	-.272 (1.174)	-.145 (1.113)	-.413 (1.061)
Concentration	-12.659* (6.569)	-14.613** (7.245)	-15.932** (6.949)
Sales growth	-.547 (.421)	-.501 (.375)	-.492 (.389)
Total patents	-2.394* (1.516)	-2.657* (1.485)	-3.051** (1.518)
Patent quality	-3.259 (2.334)	-2.656 (2.192)	-3.129 (2.310)
OIN patents	-.055 (.069)	-.050 (.064)	-.052 (.066)
SSO patents	-.021 (.071)	-.001 (.071)	-.008 (.068)
OSS demand	.047 (.099)	.156* (.093)	.158* (.096)
Year dummies	Yes	Yes	Yes
Market fixed effects	Yes	Yes	Yes
Observations	286	286	286
Log pseudolikelihood	-222.405	-219.448	-218.754

Notes: 1) Robust standard errors, clustered by market, are in parentheses. 2) * significant at 10%, ** significant at 5%, *** significant. 3) *False Commons* is measured by interacting the patent density in The Commons across markets with a time dummy that turns on for year 2003 and 2004. 4) *2003_2004* is a dummy that turns on for years 2003 and 2004. 4) *IBM patents* is measured by log of claims-weighted patent count of all IBM patents (excluding IBM patents contributed to The Commons) related to market j cumulated up to year t.

Table 8: Falsification test on the effect of The Commons on proprietary software entry, conditional fixed-effect Poisson regression

Dependent variable: Proprietary software product entry	(1)	(2)	(3)	(4)
The Commons	.062 (.054)	.051 (.074)	.131 (.096)	.122 (.101)
The Commons * Cumulativeness		.017 (.094)		0.051 (.090)
The Commons * Concentration			-.182 (.244)	-.236 (.244)
Cumulativeness	-3.566* (2.034)	-3.285 (2.890)	-3.941* (2.188)	-3.241 (2.900)
Concentration	-3.505 (6.283)	-3.671 (6.538)	-2.195 (6.593)	-2.292 (6.574)
Sales growth	.193 (.247)	.197 (.259)	.194 (.241)	.206 (.252)
Total patents	5.055* (2.796)	5.099* (2.808)	5.866* (3.076)	6.234** (3.152)
Patent quality	7.730* (5.160)	7.632* (5.277)	8.736* (5.446)	8.754* (5.514)
OIN patents	0.063 (.055)	.063* (.055)	.056 (.053)	.056 (.053)
SSO patents	-.073 (.065)	-.074 (.068)	-.080 (.066)	-.086 (.070)
OSS demand	-.204 (.191)	-.206 (.192)	-.239 (.216)	-.254 (.222)
Year dummies	Yes	Yes	Yes	Yes
Market fixed effects	Yes	Yes	Yes	Yes
Observations	232	232	232	232
Log pseudolikelihood	-443.038	-442.993	-442.390	442.066
Marginal Effects				
The Commons (average)	.062 (.059)	.062 (.054)	.090 (.059)	.100* (.056)
The Commons (cumulativeness=10%)		.054 (.064)		.078 (.066)
The Commons (cumulativeness=90%)		.072 (.082)		.131* (.081)
Test of the difference between high and low cumulativeness, p-value		.852		.571
The Commons (concentration =10%)			.105 (.069)	.119* (.067)
The Commons (concentration =90%)			. 075 (.051)	.080* (.051)
Test of the difference between high and low concentration, p-value			.457	.334

Notes: 1) 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 market, we focus only on introduction events associated with PROMT product codes. For each start-up, we only use the firm's first product in a market to capture entry. This results in 2,384 proprietary product entry events from 2002 to 2009. We also adjust for a change in the assignment of product codes during our sample. Specifically, between 2007 and 2009 application-related software products were systematically assigned to a higher product code level in the PROMT database (i.e. assigned to 7372400, Applications Software). This forced us to group several application markets together, leaving us with 29 software markets. 2) Robust standard errors, clustered by market, are in parentheses. 3) * significant at 10%, ** significant at 5%, *** significant at 1%.

Figure 1: Distribution of patents in The Commons and in IBM's portfolio by market

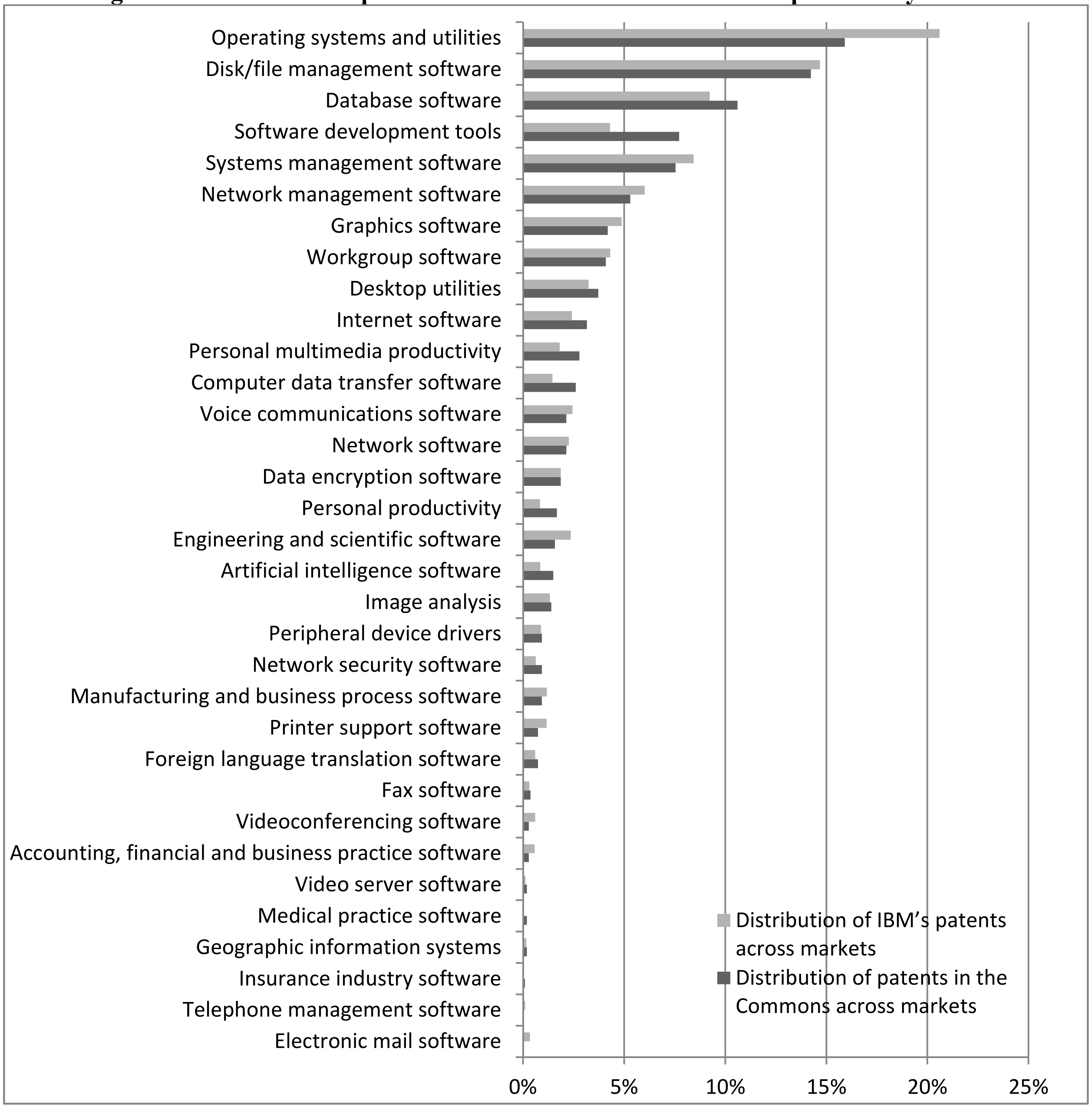
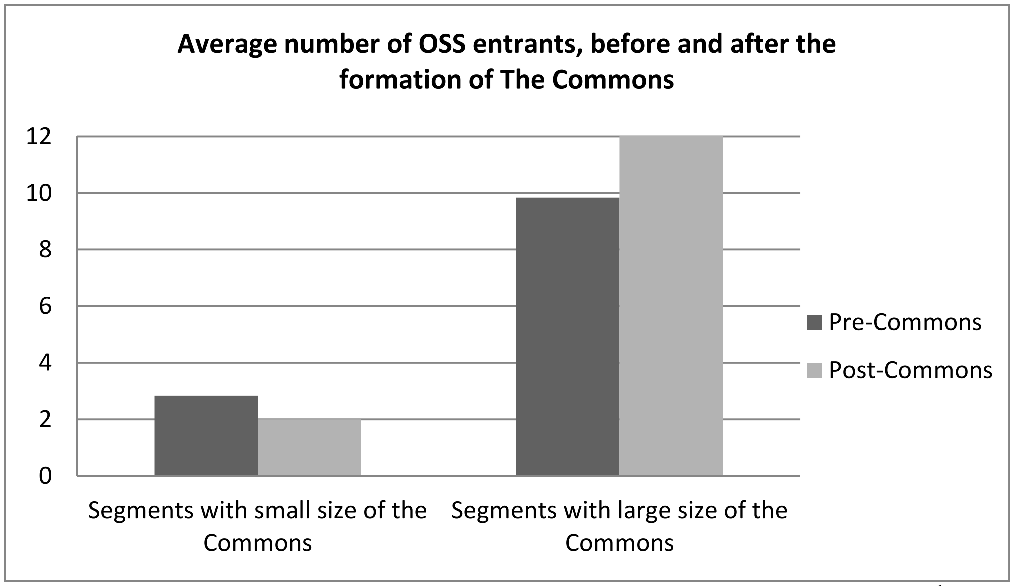
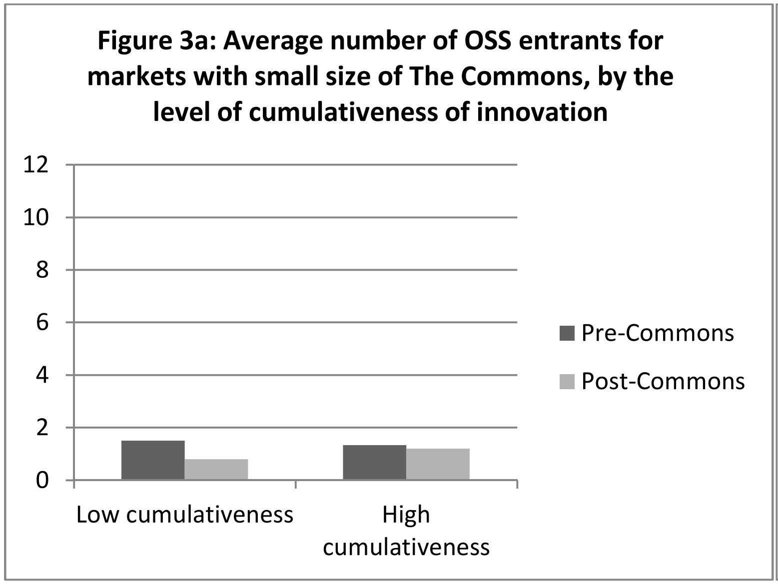


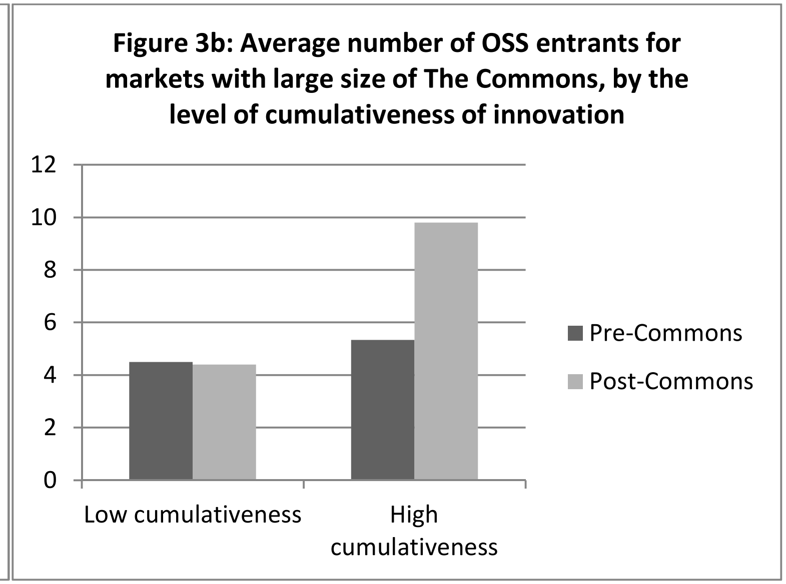
Figure 2: Average OSS entry, by the size of The Commons



Notes: 1) markets with small size of The Commons are defined as markets with bottom 25th percentile of The Commons whereas markets with large size of The Commons are defined as markets with top 25th percentile of The Commons; 2) for markets with small size of The Commons, the t-test statistic for the mean difference between pre-Commons and post-Commons OSS entrants is -.579 with a p-value 0.7; 3) for markets with large size of The Commons, the t-test statistic for the mean difference between pre-Commons and post-Commons OSS entrants is 1.313 with a p-value 0.1.

Figure 3: Average OSS entry, by the level of cumulativeness of innovation

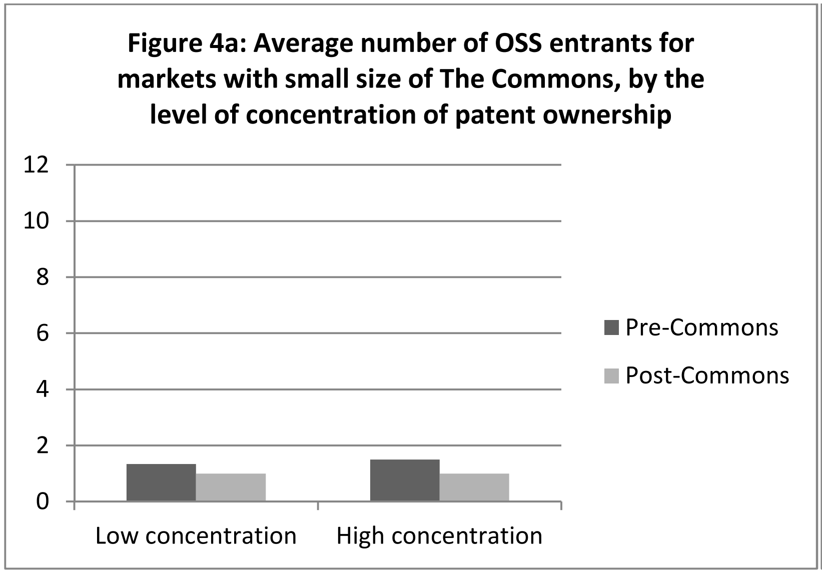


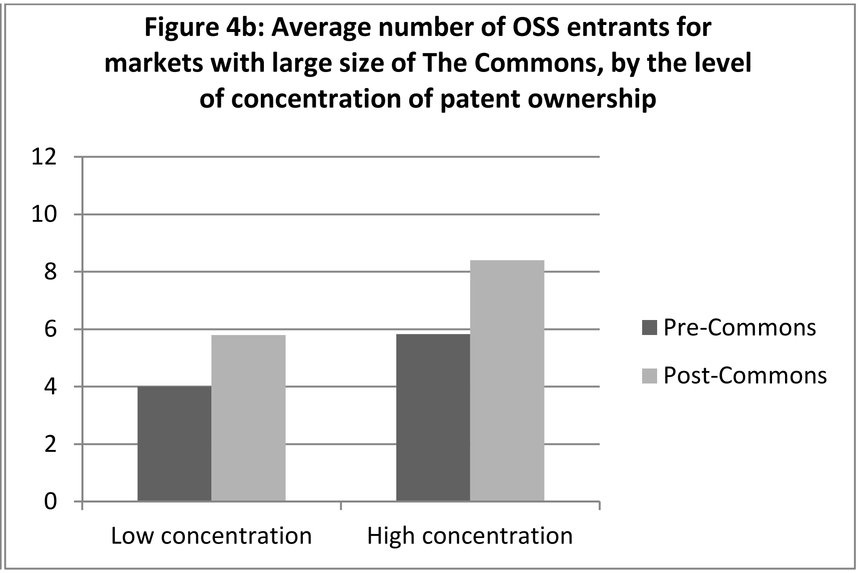


Notes: 1) markets with small size of The Commons are defined as markets with bottom 25th percentile of The Commons; 2) for markets with low cumulativeness, the t-test statistic for the mean difference between pre-Commons and post-Commons OSS entrants is -.770 with a p-value 0.8; 3) for markets with high cumulativeness, the t-test statistic for the mean difference between pre-Commons and post-Commons OSS entrants is -.232 with a p-value 0.6.

Notes: 1) markets with large size of The Commons are defined as markets with top 25th percentile of The Commons; 2) for markets with low cumulativeness, the t-test statistic for the mean difference between pre-Commons and post-Commons OSS entrants is -.054 with a p-value 0.5; 3) for markets with high cumulativeness, the t-test statistic for the mean difference between pre-Commons and post-Commons OSS entrants is 1.970 with a p-value 0.04.

Figure 4: Average OSS entry, by the level of patent ownership concentration

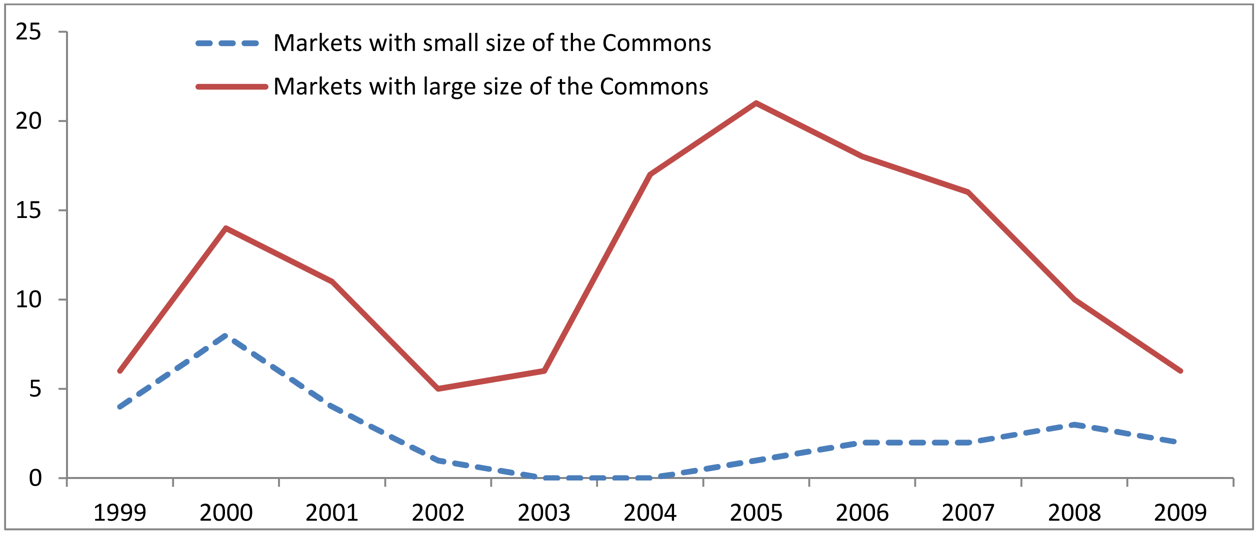




Notes: 1) markets with small size of The Commons are defined as markets with bottom 25th percentile of The Commons; 2) for markets with low concentration, the t-test statistic for the mean difference between pre-Commons and post-Commons OSS entrants is -.491 with a p-value 0.7; 3) for markets with high concentration, the t-test statistic for the mean difference between pre-Commons and post-Commons OSS entrants is -.561 with a p-value 0.7.

Notes: 1) markets with large size of The Commons are defined as markets with top 25th percentile of The Commons; 2) for markets with low concentration, the t-test statistic for the mean difference between pre-Commons and post-Commons OSS entrants is .880 with a p-value 0.2; 3) for markets with high concentration, the t-test statistic for the mean difference between pre-Commons and post-Commons OSS entrants is 1.311 with a p-value 0.1.

Figure 5: OSS entry from 1999 to 2009, by the size of The Commons



Notes: Markets with small size of The Commons are defined as markets with bottom 25th percentile of The Commons whereas markets with large size of The Commons are defined as markets with top 25th percentile of The Commons.

APPENDIX A: 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 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 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 B: Identification of Software Markets and the Matching Patent Classes

Step 1: Identify Software Markets

To measure entry with new OSS products related to different software markets in each year, a crucial step is to divide the software market into different markets that are reasonably distinct from each other. One main source of software markets 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 markets and technologies. Table B-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[37] 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 markets. Table B-2 shows some examples of the PROMT-CorpTech concordance.

Step 2: Identify Patent Classes across Software Markets

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 markets from the PROMT-CorpTech concordance, we first examined specialist firms that produce in only one software market and particularly only one CorpTech six- or seven-digit code[38]. The sample of single specialists is from the CorpTech directory, over 1992 to 2004 and 2010.[39] 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

[37] 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.

[38] Examples of CorpTech code are provided in Table B-2.

[39] 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, CM examined 28 specific product codes that have incomplete overlap with the open source product markets that we study. Finally, we took the intersection of the patent classes from the patents in The Commons with the above mapping, which lead to 34 US patent classes and their corresponding product codes.

Step 3: Match Software Markets 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 in patent class and subclass description. After identifying possible patent class-subclass, we then decided the concordance between product codes and patent class-subclass manually. This process generated the final mapping between software markets and patent class-subclass combinations. We further consolidated all product codes into 33 software markets 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 Commons helps start-ups move into new technology area. The final concordance that we used in the empirical analysis consists of 33 software markets matched to 422 patent class-subclass combinations. Table B-3 shows some examples of this final concordance between software markets and US patent class-subclass combinations. To boost our confidence, we compared our concordance with CM's concordance. Among the 28 product codes examined by CM, 18 of them closely relates to 13 software markets defined by us, though some of our software markets are broader. We compared the patent class-subclass combinations matched to the 18 product codes by CM with the patent class-subclass combinations matched to the 13 markets by us. As shown in Table B-4, we found a large degree of overlap between the two sets of mapping. Figures B-1 and B-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 B-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 B-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 B-3: Examples of the Concordance between Software Markets and US Patent Class-subclass Combinations

Software Market	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 Market 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.

Table B-4: Validation of market-patent concordance with Cockburn and MacGarvie (2011)

	Concordance by Cockburn and MacGarvie (2011)				Our concordance		
#	Software code	Code description	Patent Class	Patent subclass	Software market	Patent class	Patent subclass
1	SOF-DM-MH	Software to present data to users in a tree-like structure	707	1-10;100-104.1	Database software	707	1-10;100-102;104.1;200-206
2	SOF-DM-MR	Software to present data to users in the form of tables of rows and columns	707	1-10;100-104.1			
3	SOF-DM-Q	Database query language software	707	1-10;100-104.1			
4	SOF-OA-ME	Electronic message systems software	709	206;	Electronic mail software	709	206;
			705	008;009			
5	SOF-TS-ER	Geographic information systems software	701	2XXX	Geographic information systems	701	117-223;300
			702	005			
6	SOF-UT-H	Software to control the operation of peripheral devices	710	1-74;	Peripheral device drivers	710	8-19
7	SOF-AC-B	Invoicing and billing software	705	001-045	Accounting, financial and business practice software	705	16-45;51-59;64-80
8	SOF-AC-T	Tax preparation and reporting software	705	019;031			
9	SOF-AI-A	Voice technology software	704	<=278	Voice communications software	704	200-256;256.1-256.8;257-278
10	SOF-AI-L	Natural language software	704	8;9	Foreign language translation software	704	1-10;
11	SOF-AI-N	Software that simulates the neural pathways of the brain	706	15-44	Artificial intelligence software	706	2;10;12;14-62
12	SOF-SV-AR	Artificial intelligence R&D	706	15-62			
13	SOF-DM-F	File management software	707	1-10;200-206	Disk/file management software	711	100-173
14	SOF-MA-Q	Software to control product quality	700	108-115	Manufacturing and business process software	700	28-55;95-244;266-274
15	SOF-OA-GD	Three dimensional representation software	345	418-427	Graphics software	345	418-428;440-443;467-475;502-506;519;520;536-689
			700	98;118;119;120			
16	SOF-OA-P	Desktop publishing software	715	500-542	Personal productivity	715	5XX
17	SOF-OA-W	Word processor/text editor software	715	5XX			
			707	200-206			
18	SOF-WD-I	Inventory management software	705	28;10	Manufacturing and business process software	700	28-55;95-244;266-274

Figure B-1: Identification of Software Markets

Software Market	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 B-2: Mapping Software Markets to Patent Subclasses

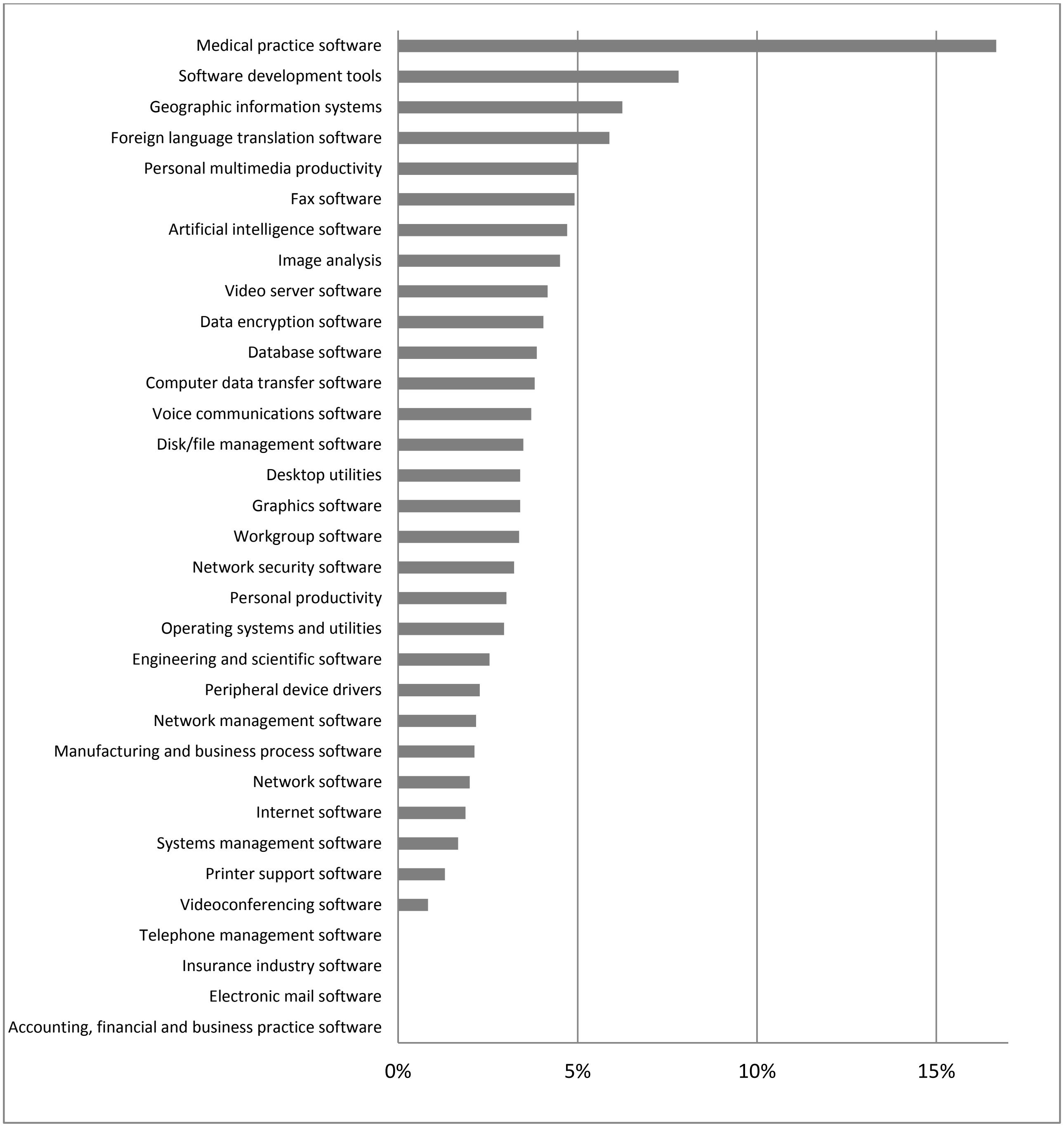
Step3

Step2

CorpTech Product Code	Patent Class	Patent Subclass
SOF-DM Keywords: Database/file management software, DBMS, Relational DBMS, Information storage and retrieval systems software (ISRS)	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	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 C: Supporting Empirical Results

Figure C-1: Fraction of IBM's patents contributed to The Commons by market



Note: As shown by this figure, the fraction of IBM patents that are contributed to The Commons are quite similar across markets with the only exception to be medical practice software market. Although it shows IBM contributed 17% of its patents in this market to The Commons, because we were only able to identify IBM's six patents related to this market by year 2004, IBM in fact only contributed one patent about medical practice software to The Commons.

Table C-1: Baseline results, linear fixed-effects regression

Dependent variable: OSS entry	Specification testing direct impact of The Commons only		Add interaction with cumulativeness of innovation		Add interaction with concentration of patent ownership		Add interaction with cumulativeness of innovation and with concentration of patent ownership	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
The Commons	.112* (.071)	.132* (.082)	.075 (.080)	.095 (.091)	-.091 (.082)	-.074 (.068)	-.121 (.091)	-.106 (.080)
The Commons * Cumulativeness			.112* (.061)	.111* (.066)			.106* (.062)	.107* (.064)
The Commons * Concentration					.724* (.405)	.730* (.384)	.708* (.396)	.719* (.375)
Cumulativeness	.381 (.672)	.583 (.780)	1.235* (.727)	1.400* (.787)	.673 (.621)	.810 (.729)	1.475** (.726)	1.591** (.791)
Concentration	.366 (3.722)	.243 (3.567)	.888 (3.570)	.742 (3.486)	1.730 (3.593)	1.596 (3.524)	2.195 (3.445)	2.053 (3.441)
Sales growth	-.479* (.316)	-.500 (.332)	-.515* (.317)	-.529* (.331)	-.552* (.354)	-.567* (.365)	-.584* (.352)	-.594* (.362)
Total patents	-.337 (.731)	-.193 (.829)	.214 (.777)	.310 (.839)	-.171 (.757)	-.098 (.798)	.346 (.752)	.384 (.775)
Patent quality	-.738 (1.179)	-.784 (1.201)	-.718 (1.210)	-.802 (1.237)	-.848 (1.166)	-.939 (1.180)	-.827 (1.198)	-.953 (1.217)
OIN patents		-.026 (.037)		-.020 (.036)		-.033 (.035)		-.027 (.034)
SSO patents		-.037 (.074)		-.037 (.071)		-.023 (.058)		-.024 (.056)
OSS demand		.003 (.015)		.009 (.016)		.012 (.023)		.018 (.022)
Year dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Market fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	363	363	363	363	363	363	363	363
R-squared	0.58	0.59	0.59	0.59	0.59	0.60	0.60	0.60
Marginal Effects								
The Commons (average)	.112* (.071)	132* (.082)	.166** (.078)	.185** (.091)	.073* (.049)	.091* (.053)	.125*** (.047)	.143*** (.055)
The Commons (cumulativeness=10%)			.100 (.077)	.119 (.087)			.062 (.054)	.079 (.058)
The Commons (cumulativeness=90%)			.253** (.101)	.271** (.116)			.207*** (.073)	.226*** (.083)
Test of the difference between high and low cumulativeness, p-value			.067	.091			.087	.094
The Commons (concentration =10%)					.003 (.046)	.021 (.041)	.057 (.048)	.073* (.047)
The Commons (concentration =90%)					.151* (.079)	.170** (.085)	.201*** (.075)	.220*** (.083)
Test of the difference between high and low concentration, p-value					.074	.057	.073	.055

Notes: 1) Robust standard errors, clustered by market, are in parentheses. 2) * significant at 10%, ** significant at 5%, *** significant at 1%. 3) R-squared includes fixed effects in R-squared computation.

Table C-2: Robustness test using raw number of patents, conditional fixed-effect Poisson regression

Dependent variable: OSS entry	Specification testing direct impact of The Commons only		Add interaction with cumulativeness of innovation		Add interaction with concentration of patent ownership		Add interaction with cumulativeness of innovation and with concentration of patent ownership	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
The Commons	.040 (.139)	.086 (.152)	-.132 (.142)	-.080 (.150)	-.411 (.265)	-.564* (.297)	-.250 (.243)	-.282 (.316)
The Commons * Cumulativeness			.686*** (.162)	.725*** (.170)			.645*** (.179)	.646*** (.208)
The Commons * Concentration					1.255** (.600)	1.824*** (.709)	.349 (.579)	.612 (.873)
Cumulativeness	1.613* (.939)	1.646 (1.254)	3.886*** (.832)	4.357*** (.983)	2.085** (.804)	2.053** (1.004)	3.890*** (.841)	4.184*** (1.146)
Concentration	-5.203 (5.298)	-3.850 (6.315)	-10.581** (4.877)	-9.905* (6.375)	-10.325* (6.728)	-7.960 (6.866)	-11.664** (5.411)	-10.644* (6.325)
Sales growth	-.439 (.342)	-.446 (.368)	-.642* (.407)	-.690* (.458)	-.500 (.359)	-.551 (.401)	-.648* (.408)	-.700 (.464)
Total patents	.663 (2.030)	-.775 (2.287)	2.113 (2.107)	2.756 (2.270)	.790 (1.928)	.870 (2.073)	2.076 (2.100)	2.574 (2.290)
Patent quality	.109 (3.232)	-.095 (3.241)	.288 (3.165)	.333 (3.130)	.541 (3.195)	.596 (3.045)	.398 (3.120)	.524 (3.040)
OIN patents		-.259 (.203)		-.339** (.168)		-.327* (.189)		-.353** (.169)
SSO patents		.038 (.197)		-.009 (.143)		.214 (.207)		.054 (.200)
OSS demand		.053 (.096)		.020 (.093)		.064 (.093)		.028 (.093)
Year dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Market fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	286	286	286	286	286	286	286	286
Log pseudolikelihood	-228.451	-227.505	-223.332	-223.024	-227.076	-225.177	-223.246	-223.835
Marginal Effects								
The Commons (average)	.040 (.139)	.086 (.152)	.377*** (.144)	.457*** (.149)	-.133 (.163)	-.158 (.177)	.306* (.185)	.333* (.226)
The Commons (cumulativeness=10%)			-.002 (.132)	.056 (.139)			-.051 (.153)	-.024 (.174)
The Commons (cumulativeness=90%)			.879*** (.223)	.988*** (.229)			.778*** (.280)	.805** (.346)
Test of the difference between high and low cumulativeness, p-value			.001	.001			.001	.001
The Commons (concentration =10%)					-.245 (.199)	-.322 (.220)	.275 (.223)	.278 (.291)
The Commons (concentration =90%)					-.005 (.136)	.029 (.147)	.342** (.152)	.395** (.166)
Test of the difference between high and low concentration, p-value					.036	.009	.547	.483

Notes: 1) Robust standard errors, clustered by market, are in parentheses. 2) * significant at 10%, ** significant at 5%, *** significant at 1%. 3) The number of observations is lower than 363 because of the use of conditional fixed effects Poisson models, which drops markets without OSS entry over the entire sample period.

Table C-3: Robustness test using raw number of patents, linear fixed-effects regression

Dependent variable: OSS entry	Specification testing direct impact of The Commons only		Add interaction with cumulativeness of innovation		Add interaction with concentration of patent ownership		Add interaction with cumulativeness of innovation and with concentration of patent ownership	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
The Commons	.136 (.126)	.176 (.140)	.005 (.131)	.045 (.139)	-.322* (.166)	-.315* (.163)	-.373* (.178)	-.351* (.171)
The Commons * Cumulativeness			.388** (.177)	.401* (.200)			.319* (.176)	.319* (.188)
The Commons * Concentration					1.633** (.786)	1.758** (.811)	1.427* (.760)	1.515* (.771)
Cumulativeness	.406 (.656)	.698 (.834)	1.957** (.784)	2.426** (1.054)	.965* (.551)	1.103 (.724)	2.173** (.808)	2.419** (.985)
Concentration	.897 (3.760)	.502 (3.681)	2.255 (3.606)	1.863 (3.583)	1.738 (3.210)	1.350 (3.320)	2.751 (3.010)	2.313 (3.172)
Sales growth	-.493* (.322)	-.544* (.347)	-.563* (.334)	-.622* (.361)	-.564* (.356)	-.612* (.377)	-.612* (.357)	-.664* (.379)
Total patents	-.132 (.976)	.212 (1.248)	1.312 (1.296)	1.832 (1.595)	.353 (1.021)	.473 (1.185)	1.482 (1.251)	1.723 (1.442)
Patent quality	-.434 (1.506)	-.347 (1.557)	.157 (1.630)	.301 (1.696)	-.217 (1.548)	-.296 (1.534)	.243 (1.617)	.212 (1.637)
OIN patents		-.204* (.111)		-.210** (.099)		-.229** (.104)		-.230** (.099)
SSO patents		-.017 (.189)		-.066 (.178)		.054 (.139)		.005 (.134)
OSS demand		-.002 (.017)		-.003 (.017)		.011 (.021)		.009 (.021)
Year dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Market fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	363	363	363	363	363	363	363	363
R-squared	0.52	0.52	0.53	0.53	0.53	0.53	0.54	0.54
Marginal Effects								
The Commons (average)	.136 (.126)	.176 (.140)	.318** (.145)	.369** (.171)	.048 (.095)	.083 (.103)	.208** (.087)	.250** (.103)
The Commons (cumulativeness=10%)			.089 (.119)	.132 (.130)			.020 (.096)	.062 (.102)
The Commons (cumulativeness=90%)			.620** (.252)	.682** (.298)			.456** (.197)	.498** (.222)
Test of the difference between high and low cumulativeness, p-value			.028	.045			.070	.090
The Commons (concentration =10%)					-.110 (.096)	-.086 (.096)	.071 (.095)	.104 (.098)
The Commons (concentration =90%)					.224 (.150)	.273* (.164)	.362*** (.137)	.413*** (.159)
Test of the difference between high and low concentration, p-value					.037	.030	.060	.049

Notes: 1) Robust standard errors, clustered by market, are in parentheses. 2) * significant at 10%, ** significant at 5%, *** significant at 1%. 3) R-squared includes fixed effects in R-squared computation.

Table C-4: Robustness test using sample with different ending year, conditional fixed-effect Poisson regression

Dependent variable: OSS entry	Sample ends at year 2008				Sample ends at year 2007			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
The Commons	.123	.025	-.162	-.126	.152	.094	-.127	-.013
	(.091)	(.133)	(.137)	(.173)	(.111)	(.155)	(.155)	(.214)
The Commons * Cumulativeness		.351***		.318***		.348***		.321***
		(.096)		(.101)		(.099)		(.109)
The Commons * Concentration			.770**	.411			.716**	.258
			(.381)	(.405)			(.361)	(.409)
Cumulativeness	1.004	3.523***	1.008	3.296***	1.030	3.436***	1.028	3.277***
	(1.244)	(1.100)	(1.113)	(1.141)	(1.194)	(1.182)	(1.100)	(1.193)
Concentration	-11.016	-16.289**	-14.951*	-17.811**	-9.434	-15.160**	-13.143*	-15.989**
	(8.088)	(8.099)	(8.432)	(7.876)	(7.568)	(7.726)	(8.012)	(7.373)
Sales growth	-.418	-.590	-.506	-.620	-.533	-.680	-.616*	-.698*
	(.398)	(.451)	(.418)	(.461)	(.354)	(.407)	(.359)	(.411)
Total patents	-.584	1.111	-1.070	.716	-.204	1.336	-.560	1.127
	(1.559)	(1.470)	(1.585)	(1.511)	(1.531)	(1.502)	(1.526)	(1.527)
Patent quality	-.851	-.645	-.820	-.631	.136	.316	.392	.397
	(2.524)	(2.437)	(2.526)	(2.433)	(2.785)	(2.437)	(2.785)	(2.666)
OIN patents	-.032	-.038	-.040	-.041	-.012	-.025	-.016	-.025
	(.071)	(.065)	(.069)	(.066)	(.080)	(.074)	(.080)	(.075)
SSO patents	-.020	-.027	.024	-.003	.013	.009	.052	.022
	(.081)	(.068)	(.086)	(.082)	(.084)	(.069)	(.087)	(.082)
OSS demand	.082	.073	.084	.074	.090	.079	.094	.081
	(.093)	(.080)	(.098)	(.084)	(.091)	(.082)	(.095)	(.084)
Year dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Market fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	260	260	260	260	225	225	225	225
Log pseudolikelihood	-212.433	-208.056	-211.016	-207.709	-188.505	-184.602	-187.404	-184.490
Marginal Effects								
The Commons (average)	.123	.294**	.013	.211*	.152	.363**	.037	.294*
	(.091)	(.132)	(.079)	(.125)	(.111)	(.163)	(.102)	(.177)
The Commons (cumulativeness=10%)		.095		.031		.166		.112
		(.128)		(.116)		(.154)		(.156)
The Commons (cumulativeness=90%)		.556***		.449***		.634***		.544**
		(.167)		(.169)		(.167)		(.231)
Test of the difference between high and low cumulativeness, p-value		.001		.001		.001		.003
The Commons (concentration =10%)			-.054	.175			-.027	.271
			(.096)	(.147)			(.118)	(.202)
The Commons (concentration =90%)			.091	.253**			.109	.320**
			(.074)	(.108)			(.094)	(.154)
Test of the difference between high and low concentration, p-value			.043	.310			.047	.527

Notes: 1) Robust standard errors, clustered by market, are in parentheses. 2) * significant at 10%, ** significant at 5%, *** significant at 1%.

Table C-5: Robustness test adding "Top 4 incumbents market share", conditional fixed-effect Poisson regression

Dependent variable: OSS entry	(1)	(2)	(3)	(4)	(5)	(6)
The Commons	-.113 (.140)	-.088 (.170)	.199 (.128)	.122 (.159)	-.069 (.148)	-.048 (.174)
The Commons * Cumulativeness		.305*** (.101)		.302*** (.097)		.294*** (.097)
The Commons * Concentration	.665* (.361)	.344 (.375)			.946** (.447)	.601 (.442)
The Commons * Top 4 incumbent market share			-.121 (.150)	-.147 (.140)	-.214 (.147)	-.188 (.143)
Cumulativeness	1.078 (1.032)	3.195*** (1.015)	1.079 (1.218)	3.153*** (.878)	1.192*** (1.044)	3.209*** (1.022)
Concentration	-10.542 (7.553)	-13.167* (6.951)			-9.879 (7.256)	-12.543* (6.726)
Sales growth	-.598 (.423)	-.716* (.408)	-.405 (.508)	-.579 (.497)	-.311 (.465)	-.441 (.467)
Total patents	-1.066 (1.578)	.624 (1.516)	-.133 (1.538)	1.570 (1.341)	.973 (1.578)	.633 (1.466)
Patent quality	-1.731 (2.422)	-1.438 (2.342)	-1.546 (2.347)	-1.190 (2.245)	-1.965 (2.418)	-1.672 (2.310)
Top 4 incumbents market share	1.294 (1.277)	1.280 (1.129)	1.483 (1.283)	1.574 (1.112)	1.649 (1.225)	1.549 (1.139)
OIN patents	-.080 (.069)	-.076 (.066)	-.069 (.071)	-.072 (.066)	-.079 (.069)	-.075 (.065)
SSO patents	.027 (.083)	.004 (.078)	.003 (.064)	.009 (.050)	.033 (.084)	.010 (.078)
OSS demand	.062 (.099)	.060 (.086)	.063 (.091)	.055 (.078)	.063 (.095)	.058 (.082)
Year dummies	Yes	Yes	Yes	Yes	Yes	Yes
Market fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	286	286	286	286	286	286
Log pseudolikelihood	-226.032	-222.856	-227.163	-223.643	-224.945	-222.031
The Commons (average)	.034 (.086)	.215* (.123)	.151* (.092)	.287** (.130)	.054 (.085)	.227* (.121)
The Commons (cumulativeness=10%)		.046 (.118)		.120 (.125)		.064 (.115)
The Commons (cumulativeness=90%)		.438*** (.162)		.508*** (.165)		.442*** (.159)
Test of the difference between high and low cumulativeness, p-value		.003		.002		.003
The Commons (concentration =10%)	-.025 (.103)	.184 (.143)			-.030 (.107)	.173 (.141)
The Commons (concentration =90%)	.102 (.079)	.250** (.108)			.150* (.079)	.288** (.112)
Test of the difference between high and low concentration, p-value	.066	.358			.034	.174
The Commons (Top 4 incumbent market share =10%)			.181* (.112)	.324** (.151)	.107 (.094)	.274** (.132)
The Commons (Top 4 incumbent market share =90%)			.118 (.084)	.248** (.115)	-.003 (.092)	.177 (.120)
Test of the difference between high and low top 4 incumbents market share, p-value			.421	.295	.148	.189

1) Robust standard errors, clustered by market, are in parentheses. 2) * significant at 10%, ** significant at 5%, *** significant at 1%.

Table C-6: Robustness test adding "Top 4 incumbents market share", linear fixed-effects regression

Dependent variable: OSS entry	(1)	(2)	(3)	(4)	(5)	(6)
The Commons	-.075 (.066)	-.106 (.078)	.112 (.097)	.072 (.112)	-.055 (.079)	-.086 (.090)
The Commons * Cumulativeness		.103* (.065)		.106* (.065)		.098 (.065)
The Commons * Concentration	.700* (.383)	.693* (.375)			.887** (.399)	.863** (.399)
The Commons * Top 4 incumbent market share			.012 (.178)	.022 (.176)	-.128 (.166)	-.116 (.167)
Cumulativeness	.828 (.733)	1.581** (.778)	.625 (.774)	1.402* (.772)	.921 (.711)	1.625** (.779)
Concentration	1.678 (3.650)	2.108 (3.533)			2.567 (3.448)	2.893 (3.380)
Sales growth	-.619* (.373)	-.637* (.373)	-.589 (.473)	-.618 (.472)	-.482 (.466)	-.512 (.467)
Total patents	-.024 (.843)	.431 (.808)	-.115 (.689)	.317 (.682)	.025 (.822)	.451 (.800)
Patent quality	-.874 (1.183)	-.898 (1.221)	-.714 (1.188)	-.757 (1.224)	-1.098 (1.292)	-1.099 (1.326)
Top 4 incumbents market share	.981 (.922)	.834 (.879)	1.366 (.846)	1.188 (.789)	1.108 (.839)	.957 (.794)
OIN patents	-.031 (.034)	-.026 (.033)	-.024 (.036)	-.019 (.035)	-.034 (.035)	-.029 (.034)
SSO patents	-.021 (.059)	-.021 (.057)	-.033 (.078)	-.034 (.076)	-.012 (.064)	-.014 (.062)
OSS demand	.011 (.023)	.017 (.023)	.002 (.017)	.009 (.018)	.009 (.022)	.014 (.022)
Year dummies	Yes	Yes	Yes	Yes	Yes	Yes
Market fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	363	363	363	363	363	363
R-square	.527	.529	.521	.523	.529	.529
The Commons (average)	.084* (.053)	.134** (.056)	.116* (.073)	.166** (.083)	.098* (.054)	.144** (.062)
The Commons (cumulativeness=10%)		.074 (.059)		.104 (.082)		.087 (.061)
The Commons (cumulativeness=90%)		.215** (.085)		.249** (.108)		.221** (.092)
Test of the difference between high and low cumulativeness, p-value		.109		.103		.135
The Commons (concentration =10%)	.017 (.041)	.068 (.048)			.012 (.047)	.061 (.054)
The Commons (concentration =90%)	.159* (.086)	.209** (.085)			.193** (.084)	.237 (.090)
Test of the difference between high and low concentration, p-value	.067	.064			.026	.031
The Commons (Top 4 incumbent market share =10%)			.114 (.081)	.161* (.092)	.126* (.068)	.170** (.077)
The Commons (Top 4 incumbent market share =90%)			.120 (.091)	.172* (.097)	.061 (.068)	.111* (.070)
Test of the difference between high and low top 4 incumbent market share, p-value			.947	.903	.440	.486

1) Robust standard errors, clustered by market, are in parentheses. 2) * significant at 10%, ** significant at 5%, *** significant at 1%.

Table C-7: GMM estimates of count data regression using each subset of IVs

Dependent variable: OSS entry	IV: IBM litigated patents				IV: IBM litigated patents X afteryear2003				IV: IBM litigated patents X afteryear2003 X Linux-related market			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
The Commons	.047 (.301)	-.031 (.208)	-1.186* (.656)	-.421 (.350)	-.094 (.390)	-.023 (.142)	-.125 (.219)	-.650* (.358)	-.051 (.256)	-.234 (.189)	-.402 (.376)	-1.053** (.410)
The Commons * Cumulativeness		.335** (.147)		.605** (.241)		.248* (.134)		.519** (.213)		.489* (.250)		.578** (.231)
The Commons * Concentration			2.806** (1.220)	1.113** (.513)			1.325* (.734)	1.865* (1.021)			1.883* (1.051)	2.618** (1.056)
Cumulativeness	-3.039 (2.989)	-2.884 (2.430)	-2.682 (2.205)	-1.011 (3.148)	-3.496 (2.884)	-2.695 (1.734)	-4.542*** (1.406)	.802 (2.593)	-2.374 (2.386)	-.793 (2.586)	-.620 (1.353)	5.172*** (1.975)
Concentration	3.396 (5.151)	.352 (4.440)	-2.460 (4.977)	-3.129 (6.019)	5.903 (5.190)	.746 (4.580)	-2.833 (5.028)	.098 (4.945)	4.170 (5.329)	-.021 (5.413)	-2.925 (5.526)	-6.148* (5.740)
Sales growth	1.670** (.718)	.847* (.551)	.204 (.484)	.041 (.643)	2.596*** (.661)	.969 (.737)	.565 (.575)	-1.132 (1.328)	3.062*** (.587)	1.203* (.675)	-.190 (.714)	-1.327** (.480)
Total patents	-.598 (1.483)	-.240 (1.248)	-.544 (1.093)	.650 (1.636)	-.334 (1.322)	.130 (.834)	-1.210* (.651)	1.711 (1.373)	.294 (1.174)	.946 (1.305)	.614 (.713)	3.677*** (1.007)
Patent quality	2.551 (2.046)	4.112*** (.923)	3.011*** (1.128)	3.756*** (.975)	3.870* (2.044)	4.547*** (1.030)	3.396*** (1.032)	2.410* (1.390)	3.271** (1.624)	4.737*** (0.909)	2.412 (1.045)	1.605 (1.279)
Year dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	363	363	363	363	363	363	363	363	363	363	363	363
Over-identification, p-value	0.38	0.16	0.53	0.13	0.96	0.17	0.17	0.48	0.59	0.14	0.12	0.13

Notes: 1) The full set of IVs include: IBM litigated patents, IBM litigated patents X afteryear2003, and IBM litigated patents X afteryear2003 X Linux-related market. 2) Robust standard errors, clustered by market, are in parentheses. 3) * significant at 10%, ** significant at 5%, *** significant at 1%. 4) We use Wooldridge's quasi-differencing transformation to remove market fixed effects.