Research Policy 51 (2022) 104477

Contents lists available at ScienceDirect

Research Policy

journal homepage: www.elsevier.com/locate/respol

Profiting from digital innovation: Patents, copyright and performance

Kevin J. Boudreau [a,1], Lars Bo Jeppesen [b,1], Milan Miric [c,*,1]

[a] Northeastern University, United States of America
[b] Copenhagen Business School, Denmark
[c] University of Southern California, United States of America

A R T I C L E I N F O

Keywords:
Copyright
Patents
Digitization
Digital platforms
Mobile apps
Platform complementers
Innovation

A B S T R A C T

It is yet unclear whether patents and copyright are effective at protecting digital innovations. In this paper, we investigate this question using novel product-level data on mobile apps, in which we relate the use of both patents and copyright to (i) revenue performance and (ii) IP licensing. We theorize that these relationships depend on differences in product-level characteristics and that apps differentiated by their design are more likely and effectively to be protected by patents; apps combining elements of differentiated content are more likely and effectively protected by copyright. Our results support these predictions that product characteristics shape the appropriate contingent use of patent and copyright protection in digital products. These patterns are especially relevant to industries where digital products combine elements of differentiated design and differentiated digital content.

1. Introduction

An important strategic issue for organizations is protecting their innovations so that they may capture value and profit from their innovations (Teece, 1986; Cohen et al., 2000). As digital innovation has grown in prominence, studies have explored how the strategies used to protect digital products and services may differ from the non-digital settings that are typically studied in this literature (Luo and Mortimer, 2017; Nagaraj, 2018; Cockburn and MacGarvie, 2009, 2011; Hall and Ziedonis, 2001; Huang et al., 2012). The interest in protecting digital innovations is, in part, motivated by the fact that patents and copyright, which are two of the most common formal protection strategies, might not be particularly effective at protecting computer code (i.e., machine instructions), which is often the basis for any digital product (Goldfarb et al., 2014; Varian, 2005). This raises the question of when patents and copyright might be effective at protecting digital products, which types of innovations they best protect, and whether they can be effective.

Earlier studies looking into the use of patents and copyright typically explained the use of these strategies based on industry-level characteristics, such as the effectiveness of legal protections or industry complexity (Teece, 1986; Cohen et al., 2000). Software and digital industries have features which could make patents and copyright less effective at protecting digital innovations. For example, many digital innovations are based on code, which may be easily reinvented or replicated and invented around even when protected by formal Intellectual Property (IP) rights. Yet, at the same time, there is evidence that some firms do use patents and copyright to protect digital innovations (Graham et al., 2009; Miric et al., 2019). Therefore, it remains unclear whether these strategies are effective. For instance, there is evidence that while patents may reduce competitive pressure and entry (Cockburn and MacGarvie, 2011), but at the same time only 13% of digital companies use patents (Miric et al., 2019).

In this paper, we study whether and when patents and copyrights are effective at protecting digital innovations, meaning whether their use is associated with higher revenues (or sales) when protecting certain types of products. We argue that the use and effectiveness of patents and copyright protections is related to the specific ways in which products are differentiated.

Products that are novel or distinct may be effectively protected by patents, as patents are intended to protect technological innovations. By contrast, copyright provides only limited protection for technological innovations, as imitators may be able to replicate the functionality without copying the code (the only protection that copyright provides). Alternatively, products that are differentiated on the basis of content would not be protected by patents as patents provided limited protection for specific expressions (text, images, or videos). Copyright, by contrast, is intended to protect specific expressions. In addition to these products being more likely to be protected by patents or copyright, we also expect that corresponding use of patents or copyright in these cases

* Corresponding author.
E-mail address: mmiric@marshall.usc.edu (M. Miric).
[1] All authors contributed equally to this paper.

https://doi.org/10.1016/j.respol.2022.104477
Received 1 February 2021; Received in revised form 3 December 2021; Accepted 8 January 2022

will be associated with higher revenues, which would suggest these protections are effective at preventing imitation.

We test these predictions using data on mobile software application developers on the Apple App Store. Mobile apps are representative of many modern digital products. We combine observational data on products and companies in the marketplace with survey-based data, which allows us to capture both protection (patent and copyright) and product design strategies.[2] Our dataset covers 1554 products by 581 developers. The observational data allow us to compare our survey sample to the population of companies.

First, we explore whether the degree to which a product is differentiated on the basis of design and the degree to which a product is differentiated by the content it embodies is associated with their use of patents and registered copyright. We find that developer firms whose products are differentiated on the basis of design are 17% more likely to use patents; but, are not more likely to be associated with the use of copyright, respectively. Conversely, developer firms whose products embody differentiated content are 29% more likely to use registered copyright; but are not more likely to be associated with greater use of patents.

Second, we test whether using patents to protect products that are differentiated on the basis of design and using copyright to protect products that are differentiated on the basis of content are associated with higher revenues. Given that property rights are intended to limit competition, observing higher revenues when patents and copyright are used would be consistent with these being effective protections (Teece, 1986). We find that patent use in combination with design differentiation is associated with 44% higher revenues than baseline levels (we find no such relationship with revenues where patents are used with products differentiated on the basis of content). Registered copyright use is associated with 23% higher revenues for products embodying differentiated content relative to baseline levels (we find no such relationship with revenues where registered copyright are used with products differentiated on the basis of design).

Third, we consider whether products protected by patents or copyright are more likely to be licensed to other companies. If patents and copyrights are effective at protecting certain innovations, we would then expect that they would be associated with technology transfer or licensing (Arora et al., 2004). We find that the likelihood of licensing is 33% higher for firms that use patents and protect products differentiated on the basis of design. The analysis is robust to a variety of specifications and econometric tests, including coarsened exact matched (CEM) sample to account for unobserved quality differences between firms that employ design or content strategies. We also look at non-effects for relationships that are not consistent with our predictions (i.e., copyright and design) to check for omitted variable bias.

This paper makes a number of contributions to the literature. First, while there is considerable research looking at how industry characteristics shape the use of patents and copyright (Teece, 1986; Cohen et al., 2000), there has been less work looking at variation in product characteristics and IP strategies within the same industry. In this paper, we provide evidence regarding how product-level characteristics shape the use of patents and copyright within digital industries. This allows us to explain some of the heterogeneity that can be observed in patent and copyright use in digital industries, explaining why only certain digital products are protected by patents, and others by digital copyright. The product-level focus also differentiates this paper from studies that have focused on firm-level characteristics such as firm size or funding sources in determining the choice of protection strategies (Miric et al., 2019; Leiponen and Byma, 2009; Graham et al., 2009). Second, as mentioned, existing studies have looked at the role of using patents to protect software, which is often conceptualized as a product comprised only of "code" (Graham et al., 2009; Cockburn and MacGarvie, 2011; Lerner and Zhu, 2007; Bessen and Hunt, 2007). However, as we discuss, protection strategies such as copyright may not effectively protect code from imitation, but instead may be used to protect other aspects embedded in digital products. Finally, our study adds to recent research considering registered copyright as a protection strategy (Nagaraj, 2018; Luo and Mortimer, 2017; Lerner and Zhu, 2007). This highlights the importance of considering both patent and registered copyright when studying protection strategies for digital products, as both are relevant protection strategies, but for different types of multi-faceted products. All of the results of this paper speak to the broader conversation regarding how companies may use property rights to insulate themselves from competition (Teece, 1986; Cohen et al., 2000) by extending this to the context of digital industries and specifically protection strategies for complementers in digital platforms.

2. Background and related literature

Intellectual Property (IP) rights can play an especially important role in facilitating innovation by providing innovators with exclusive rights to use their creation, either to commercialize the innovation themselves or license it to others (Teece, 1986; Cohen et al., 2000). The institutions currently governing IP rights were established centuries ago and they were not designed to protect digital products.[3] For instance, patents were originally intended to protect mechanical creations, while copyright was intended to protect books and written works. These property rights have only gradually evolved, whereas the industries and technologies they are intended to protect have changed quite rapidly (Contigiani et al., 2018; Fosfuri et al., 2008). Perhaps the most prominent example is the growth of digital products, which are not neatly protected by patents and copyright and raise the question of how analog IP rights apply in a digital world (Lessig, 2008).

There are several reasons why patents and copyright rights might not be effective at protecting digital innovations. For instance, the ease of replication, reverse engineering, and re-use inherent in digital technologies imply that it is difficult to prevent imitation (Varian, 2005; Landes and Posner, 2003). An example of this is copyright protection for software. While registered copyright protection can prevent competitors from replicating the code embedded within a software application, it does not prevent them from reverse engineering and recreating that functionality themselves (Samuelson and Scotchmer, 2002). Additional challenges arise, given that many digital innovations are distributed digitally across the internet, which often cross legal jurisdictions and allows anonymity, making enforcement of property rights challenging (Waldfogel, 2012; Waldfogel and Aguiar, 2018). Additionally, the patentability of software has shifted over time, raising questions regarding its applicability and usefulness (Lerner and Zhu, 2007). Existing research does not clearly explain whether or when property rights might effectively protect digital innovations and provides potentially contradictory results.

Existing studies that have looked specifically at the effectiveness of IP rights in digital settings provide mixed evidence (Hall and Ziedonis,

[2] When validating the data we collected, we found that this method provides a more reliable measure of property rights usage, such as patents and copyright, than observational data from patent databases (e.g. USPTO). This is, in part, because many of these companies are very small, and their assignees are often the founder rather than the company making them difficult to match.

[3] In the United States, the establishment of a national system of patents and copyright was concurrent with the founding of the Constitution, with the first Patent Act drafted in 1790. Leading areas for patenting were mechanics, materials processing and handling, agriculture, animal husbandry, food, and heating (Marco et al., 2015). Copyright was primarily applied to books, translations, and derivative works, at a time when the means of producing and replicating these items had not yet been widely diffused.

2001; Cohen and Lemley, 2001; Samuelson, 1993; Graham et al., 2009). Various studies have examined which form of IP – patent, copyright or trademark –, best applies to digital content and have gone on to analyze IP rights individually (Graham et al., 2009; Lerner and Zhu, 2007). Several empirical studies have documented associations that suggest patents may be effective at protecting digital innovations, considering factors such as entry deterrence (Cockburn and MacGarvie, 2011, 2009), market valuation, survival rates (Wagner and Cockburn, 2010), and sales volumes (Lerner and Zhu, 2007).

In seeming contrast to findings that suggest patents can be effective, software patents appear to be used infrequently. Bessen and Hunt (2007) find patents are used by just 5% of software publishers. Remarkably, 75% of software patents are obtained by companies in manufacturing industries (e.g., chemical, computer, electronics, and instrument industries), with only 25% obtained by specialized software firms. Examining data from the Berkeley Patent survey of early-stage technology companies, Graham et al. (2009) find that executives of software companies regard patents to be far less important than those in other industries to gaining a competitive advantage. Furthermore, patents are reported to provide weak incentives to innovate. Even among specialized software publishers using patents, Bessen (2003) finds they often use patents for strategic purposes rather than for the enforcement of IP rights. Hall and MacGarvie (2010) find no significant stock market valuation effects associated with patents for pure software firms. With regard to copyright, as noted in the introduction, its prevalence has diminished in software since the 1980s (Lemley et al., 2006).

There are also mixed results in relation to digital products beyond software. Many studies begin with the presumption of weak IP rights rather than directly testing the point. For example, a growing stream of research on digitally recorded music tests implications for both product demand (Waldfogel, 2012; Oberholzer et al., 2015; Givon et al., 1995; Reimers, 2016; Nagaraj, 2018) and the supply-side provision of products (e.g., Mortimer et al., 2012; Givon et al., 1995; Peitz and Waelbroeck, 2006; Waldfogel and Aguiar, 2018).[4] A handful of other studies more directly examine the effects of applying or removing property rights protections and find that they are to some extent effective. For example, Nagaraj (2018) shows that copyright protection of the magazine *Baseball Digest* (vs. its out-of-copyright issues) led to 135% fewer images being shared and reused on Wikipedia and a readership reduction of the copyrighted pages of 20%. Heald (2014) finds that copyrighted music is less frequently used in movies than non-copyrighted music.

In summary, empirical research on digital IP rights provides mixed findings on their use and effectiveness. Relevant studies have conducted extensive examinations of how IP rights have been used to protect software (Hall and Ziedonis, 2001; Cohen and Lemley, 2001; Samuelson, 1993; Graham et al., 2009). These studies have generally sought to determine which form of IP rights – patents, copyright, or trademarks – best applies to digital content (Lerner and Zhu, 2007; Graham et al., 2009). The majority of these studies have sought to establish, at the level of an industry or setting, whether patents and copyright are effective means of protecting digital products. A smaller number of studies have looked at the use of patents and registered copyright at the firm level considers which types of companies choose to use patents and register copyright. However, there has been limited research into patent and copyright use within an industry, considering how product characteristics determine whether a product is protected by patent or registered copyright. While we might expect that higher-quality or more novel products might naturally be more likely to be protected, we may also expect that product characteristics could determine whether patents or registered copyright are used to protect them. These product characteristics may explain why patents and copyright are used to protect some products but not others.

3. Hypothesis development

Patents are intended to protect inventions that are novel and non-obvious. In the past, pure software such as computer code (machine instructions) was often not directly patentable because it was seen as analogous to algorithms or mathematical formulas, which cannot be patented even if they are novel and non-obvious. However, a series of legal decisions made it possible to file and enforce patents in order to protect pure software instructions (Lemley et al., 2006). Provided that a digital product has novel or inventive elements, its creators can apply for patent protection. Products that have more novel or innovative elements are more likely to be protected by patents. It is important to acknowledge that some products that are not particularly novel might also be protected by patents. However, we would expect that products that are more differentiated are more likely to be patented.

Additionally, given that patents are designed to exclude or limit competition, we would then expect that when patents are used they are associated with higher revenues (Ceccagnoli, 2009; Arora et al., 2008). We would expect this to be greater for products that are based on greater differentiation, as these products are more distinct from those already in the market and patents provide a greater breadth of protection. By contrast, products that are less similar to others in the market may gain some protection, but would not be able to drastically limit competition because many competitors might already be in the market.

Finally, effective IP rights enable easier technology transfer and licensing (Gans and Stern, 2003; Fosfuri et al., 2008), as having clear ownership over a technology makes it easy to contract and trade that technology. Therefore, much like in the arguments regarding higher revenues above, we would expect that using patents may be associated with a greater likelihood of licensing, particularly in cases where products are differentiated based on design such that they are more distinct from those already in the marketplace.

These arguments form the basis for the following hypotheses:

***Hyp* 1.** *Products differentiated on novel design are associated with greater patent use.*

***Hyp* 2.** *Products differentiated on novel design and protected by patents are associated with higher revenues.*

***Hyp* 3.** *Products differentiated on novel design and protected by patents are more likely to be licensed to other companies.*

Copyright is designed to protect the particular expression of an idea, such as in original text, images, or video. By default, copyright exists for any original expression (any text, image, or code created). Having a copyright registered makes it easier for an individual to demonstrate ownership of a particular creation and for the copyright to be enforced.[5] Software and code, specifically, are protected by copyright in the same way as any original text. However, these protections for software are suspected to be fairly weak because of the ease of both reverse engineering and inventing around such software instructions (i.e. recreating that functionality by writing new code) (Samuelson and Scotchmer, 2002). Even in cases where imitators had to violate copyright in order to invent around a software product, the courts found that copyright

[4] This work on weak IP rights in recorded music also tends to focus on consumer piracy rather than involuntary spillovers to competitors. In digital stock photos, work by Luo and Mortimer (2017) begins with a similar presumption of widespread piracy, with tens of thousands of copyright violations observed. The authors proceed to show, through an experiment, that violators are more likely to accede to copyright by informal requests rather than formal legal requests, threats, and price-related factors (Luo and Mortimer, 2017).

[5] This is the case for copyright protection in the United States.

could not be used to limit competition because that was the domain of patents (Lemley, 1995). Therefore, patents are likely not effective at protecting software even though it may be differentiated by novel designs, as argued in the case of patents above.

That said, copyright protection for other types of expression, such as video, images, and user-generated content, is more effective at preventing imitation because the content may be more difficult to invent around (Nagaraj, 2018; Reimers, 2016; Luo and Mortimer, 2017). When code is combined with other digital content such as images, video, and user-generated content, registered copyright provides a more effective strategy for protecting digital innovations. Many digital products are not created not only with software applications but also often combined with other digital content such as images and videos. Therefore, the more that products are "mashups" of code with content such as images, the more likely they are to be protected by copyright. This is likely to translate into greater effectiveness in terms of greater revenues associated with copyright use in the case of products differentiated by content. Additionally, we are likely to observe a greater likelihood of licensing in the case where products are differentiated on the basis of content while being protected by copyright, as argued above.

***Hyp* 4.** *Products differentiated on unique content are associated with greater copyright use.*

***Hyp* 5.** *Products differentiated on unique content and protected by copyright are associated with higher revenues.*

***Hyp* 6.** *Products differentiated on unique content and protected by copyright are more likely to be licensed to other companies.*

4. Dataset

The data studied in this paper cover fine-grained information, covering 581 app developers and 1554 apps on the Apple App Store. We collected data from two sources. First, we machine-collected observational data on the full population of app developers in 2013. These data provided information about developers and their products, including prices, release dates, descriptions, and other characteristics. This also provided us with support or contact email addresses that we used to construct a sample of firms to be surveyed. The sampling and response of developers who provided email addresses were likely not random, a point we address in later sections. Using this contact information, we surveyed developers and matched respondents to observation-level data, yielding a dataset covering patent and copyright use, product-level revenues, licensing and outsourcing, and product and developer characteristics.

4.1. Variable construction

IP and copyright use. We use survey responses to construct dummy variables, $PATENT$ and $COPYRIGHT$, which indicate developers who use patents and registered copyright, respectively. We do not distinguish among different possible classes of patents, although we do distinguish whether copyright had been formally registered with the U.S. Copyright Office (i.e., rather than the default coverage provided by copyright law without explicit registration).[6]

Measuring product differentiation: Design and content. We measure product characteristics based on survey responses that indicated that some aspect of the developer's products possessed an element of novel design or proprietary content.[7] We constructed two measures: one to reflect the level of differentiation based on design (*DESIGN*) and the other to reflect the level of differentiation based on content (*CONTENT*). The measures were constructed by asking developers how they intended to distinguish their products from those of competitors and by presenting a list of options and corresponding boxes that they could check in the affirmative.[8]

The variable *DESIGN* sums the number of affirmative responses in relation to the following question: Did you create *a different or novel type of app*, *pioneer new types of apps*? or *engage in rapid innovation?* and*engage in versioning*?. The variable *CONTENT* sums the number of affirmative responses in relation to the following question: Did you *use specialized content?* Did you*use user-generated content*? While these measures are imperfect, they do capture coarse but meaningful variation in the degree of differentiation each of these dimensions.

Revenues. To study the effectiveness of patents and copyrights, we analyze the revenues generated by developers from selling their products. Revenues are obtained using ranking charts and imputing revenues using the method described by Garg and Telang (2013). These revenue estimates reflect the amount of revenue generated by a particular application through both sales and in-app purchases.[9] The logarithm of total sales (USD) generated by an application is indicated by the variable $REVENUES$.[10]

Licensing and IP trade. To measure whether developers are engaging in licensing, we define the variable *IP_TRADE* as an indicator variable in cases where the developer engages in either the licensing or outright sale of apps and app technologies to other firms (rather than selling apps directly to consumers). This measure was taken from the survey in which we asked firms about their sources of revenues.

Developer and product characteristics. Using observational (app store market) data, we construct a number of control variables : the market niche of each individual app (a series of 27 indicator variables corresponding to app categories described on the App Store[11]), the average quality rating of each product (based five-point user rating scale), and entry timing (first appearance date). From the survey, we observe firm size as a categorical variable (part-time, 1 employee, 2 employees, 3–5 employees, 11–20 employees, 21–50 employees, or more than 50 employees), an indication of whether the developer is

[6] We initially attempted to use public records of patents and copyright data but found the matching of filings to individual suppliers and ensuring the filings are related to apps (rather than some other part of the business or multi-unit corporation) to be unreliable.

[7] We considered a number of approaches to measuring the different types of digital elements contained in apps that could be considered innovative, including text-based analysis and expert judgment. We chose a survey-based measure in which we directly asked the supplier whether and how they intended to differentiate their product. This approach had the advantage of most accurately capturing the intended and expected basis of innovative differentiation while drawing on the developers' own detailed technical knowledge of the design.

[8] The order in which the options were presented was randomized, and there was no compulsion to respond in the affirmative to any option.

[9] We do not capture the amount of revenue generated through advertising, although we control for advertising revenue models in our econometric specification. In the time sample studied, advertising revenue models were not particularly lucrative, so they do not have a significant influence on our empirical results. Additionally, our revenue estimates do not capture revenues generated from licensing technologies to other companies.

[10] Our measure of REVENUES (total sales USD) is based on the entire lifetime that these products were in the marketplace and does not suffer from truncation issues because of the long time series available.

[11] The categories are books, business, catalogs, developer tools, education, entertainment, finance, food and drink, games, graphics and design, health and fitness, healthcare, lifestyle, medical, music, navigation, news, photo and video, photography, productivity, reference, social networking, sports, travel, utilities, video, and weather.

Table 1
Variable Definitions & Descriptive Statistics.

Variable	Description	Mean	S. D.	Min.	Max.
Outcome Variables					
PATENT	Developer reports using patents as a way of protecting their technology.	0.11	0.31	0	1
COPYRIGHT	Developer reports using copyrights as a way of protecting their technology.	0.17	0.38	0	1
REVENUE	Total revenues generated from Product Sales (Sales and In-App Purchases). This does not suffer from right truncation because of long time series of the data. Excludes revenues from advertising or donations which we accounted for with control variables. Values log transformed.	1.19	1.24	0	7.46
LICENSING	Developer reports attempting to license their technology to other companies.	0.10	0.31	0	1
Expanatory Variables					
DESIGN (*Differentiation*)	Design differentiation is defined as the number of differentiation strategies that a developer reported using. Individual differentiation strategies are captured from the survey and listed below.	1.03	0.94	0	3
-Different Types of Apps		0.38	0.49	0	1
-Pioneering New Types of Apps		0.30	0.46	0	1
-Versioning of New Apps		0.39	0.49	0	1
CONTENT (*Differentiation*)	Content differentiation is defined as the number of content differentiation strategies that a developer reported using. Individual differentiation strategies are captured from the survey and listed in each row.	0.36	0.60	0	2
-Specialized Content		0.20	0.40	0	1
-User Generated Content		0.17	0.38	0	1
Controls & Additional Variables (E.G. IV, etc.)					
Market Tenre	Month when the developer entered (first appeared in) the marketplace. Calculated as the number of months since the beginning of the dataset.	25.62	13.66	0	58
Quality Rating	Average Rating of a Developers Products on the Apple App Store.	3.47	0.91	1	5
FIRM SIZE	Firm size based on the number of employees reported in the survey.				
<1 Employee		0.20	0.40	0	1
1 Employee		0.23	0.42	0	1
2 Employee		0.13	0.34	0	1
3–10 Employees		0.30	0.46	0	1
11–50 Employees		0.12	0.32	0	1
> 50 Employees		0.03	0.16	0	1
Apps Contain Advertising	Developer generates revenues from advertizing embedded in their applications.	0.28	0.45	0	1
Developer Creates Apps for Others	Developer generates revenues by creating applications for other companies or individuals.	0.37	0.45	0	1
Monetary Motivation	Developer reports attempting to generate revenue from the sale of their products.	0.75	0.43	0	1
Hobbyist Motivation	Developer reports being motivated to create software as a hobby. This is not the opposite of the income variable. There are some that include a positive response on both.	0.25	0.43	0	1

an amateur hobbyist developer, and a series of indicator variables related to how the developer derives income (i.e., downloads, in-app purchases, advertising, or as a complimentary service or good).

We present descriptive statistics and the correlation table for these variables in Tables 1 and 2.

4.2. Representativeness of survey sample

Survey data collection defines the sample available for the analysis, as observational data from the platform cover the entire population. The survey data were generated by, first, machine-crawling the entire U.S. Apple App Store in 2013, covering 253,100 app developers. Of these, a subset of 23,048 developers listed a contact email address in their product descriptions on the App Store. Of the recipients, 809 developers of 7973 products provided completed responses. We excluded those companies that were not attempting to generate revenue from the sale of their products directly on the App Store. These include airlines, chain restaurants, service providers, etc. all of whom may be lucrative businesses but are not trying to sell apps or games to consumers.

Despite the relatively large number of respondents compared to past survey-based studies of IP rights,[12] the sample is small in relation to the population. Anticipating this outcome when designing the research approach, we selected a design that did not rely on random sampling and allowed us to evaluate the degree of non-random sampling. Our population-level data contain all the firms in the marketplace, along with rich product- and developer-level data. Using this information, we were able to compare the distribution of the sample and population for observable characteristics to analyze the bias in our sample.

In Fig. 1, we present the distribution of key variables observable across the population of developers. Gray indicates the distribution of observations in the Population. Gray indicates the distribution of observations in the sample. PERFORMANCE is the total days that an app stays in the within-category top 10 list. The observations in our sample perform slightly better on average than those in the population (specifically, we under-sample products that have very poor performance, as measured by very few or no days in the top ranking). ENTRY COHORT is the year of first entry into the marketplace. The

[12] The number of respondents is similar to those of Yale and Carnegie surveys.

Table 2
Pairwise Correlations of Main Variables.

	Variables	1	2	3	4	5	6	7	8	9	10	11	12	13	14
	Dependent Variables														
1	*PATENT*	1.00													
2	*COPYRIGHT*	0.35	1.00												
3	*REVENUE*	0.11	−0.03	1.00											
	Developer Level Covariates														
4	*DESIGN (Diff.)*	0.17	0.16	−0.16	1.00										
5	*CONTENT (Diff.)*	0.14	0.14	0.03	0.34	1.00									
6	*Market Tenure*	0.09	0.20	−0.26	0.08	0.06	1.00								
	Firm Size														
7	< 1 Employee	−0.13	−0.13	0.07	−0.23	−0.11	−0.06	1.00							
8	1 Employee	−0.14	−0.12	0.00	−0.03	−0.12	−0.14	−0.18	1.00						
9	2 Employees	−0.09	−0.13	−0.05	−0.11	−0.06	0.02	−0.14	−0.19	1.00					
10	3–10 Employees	−0.02	0.00	0.07	−0.03	−0.01	0.04	−0.25	−0.34	−0.26	1.00				
11	11–50 Employees	0.24	0.27	−0.13	0.28	0.27	0.09	−0.18	−0.25	−0.19	−0.35	1.00			
12	> 50 Employees	0.17	0.16	−0.01	0.11	0.00	0.08	−0.07	−0.09	−0.07	−0.13	−0.09	1.00		
13	*LICENSING*	0.36	0.30	0.02	0.18	0.19	0.22	−0.12	−0.13	−0.14	−0.02	0.29	0.18	1.00	
	Product Level Covariates														
14	*QUALITY RATING*	0.03	0.09	−0.05	−0.01	0.00	0.29	−0.01	−0.03	−0.02	0.06	−0.03	0.02	0.09	1.00

Note. 27 Product Categories not reported here.
Other variables used in the analysis are described in the text.

Table 3
Comparison of Sample and Population Characteristics.

Variable	Sample			Population		
	Mean	*S.E.*	*Std. Dev.*	*Mean*	*S.E.*	*Std. Dev.*
QUALITY RATING	3.46	0.02	1.14	3.53	0.00	1.22
TOTAL PRODUCTS	11.48	0.75	20.64	4.40	0.04	17.47
PRICE	2.62	0.27	24.47	1.65	0.01	10.50
ENTRY COHORT	2010.86	0.02	1.23	2011.29	0.00	1.20
TOTAL RATINGS	2.53	0.03	2.16	2.15	0.00	1.99
PERFORMANCE	2.31	0.26	1.83	2.05	0.16	1.43

sample and population map closely, with the exception that the sample slightly favors companies that have been present in the market for some time. QUALITY RATING is the average product rating (out of five) given by customers. TOTAL RATINGS is the number of product ratings given by customers, which can also be interpreted as an approximation for downloads. PRICE is the average price charged for products at the developer level. The distribution of the sample and population overlap quite closely. TOTAL PRODUCTS is the number of different titles launched by the developer. There is a much higher share of single-product firms in the population than in the sample, which contains firms that have a larger portfolio. These comparisons suggest that we oversample larger firms. Larger firms may be more likely to use IP rights but are also more economically important and therefore more interesting to study. We also tested the robustness of our results using sampling weights (Manski and Lerman, 1977) to correct for this type of sampling bias and found consistent results (see Table 3).

5. Results

Here, we report the results of our analysis. First, we test whether products differentiated based on design or content are associated with greater use of patent and registered copyright, respectively. Second, as a way of looking for further evidence to support whether patents and copyrights are effective at protecting these types of products, we confirm whether the use of patents to protect designs is associated with higher revenues and greater likelihood to engage in technology licensing, both of which indicate that patents are in fact used to protect such products. We also test for similar patterns for the use of registered copyright and unique content. Finally, we validate these results with a variety of statistical checks and alternative specifications.

5.1. Use of IP rights by developers

5.1.1. Variable construction and analysis

In the first part of the analysis, the aim is to provide an empirical test of **Hyp 1** and **Hyp 4**. We study how the use of IP (*PATENT*, *COPYRIGHT*) correlates with the level of differentiation (*DESIGN*, *CONTENT*). We control for firm characteristics (**DEVELOPER**) and other product characteristics (**PRODUCT**). We estimate the effects for patents and copyright use jointly to account for potential correlation between the two IP choices. The basic model used in the analysis is represented below as a linear expression.[13]

$$Pr(PATENT) = \alpha_p + \beta_{p,d} DESIGN + \beta_{p,c} CONTENT + \mathbf{PRODUCT}\delta_p + \mathbf{DEVELOPER}\theta_p + \varepsilon \quad (1)$$

$$Pr(COPYRIGHT) = \alpha_c + \beta_{c,d} DESIGN + \beta_{c,c} CONTENT + \mathbf{PRODUCT}\delta_c + \mathbf{DEVELOPER}\theta_c + \varepsilon \quad (2)$$

where α and β are coefficients, δ and θ are vectors of coefficients for control variables, and ε refers to zero-mean error terms to be estimated. The unit of analysis is at the developer (firm) level.[14] We refer to our coefficients using the subscript $\beta_{IP\,Type,\,Differentiation}$. For instance, the coefficient for the impact of $DESIGN$ on the use of patents is denoted as $\beta_{p,d}$. Based on earlier arguments, coefficients $\beta_{p,d}$ and $\beta_{c,c}$ are predicted to be positive. However, we do not expect to find any effect for other design choices, $\beta_{p,c}$ and $\beta_{c,d}$. The two model expressions

[13] In our analysis, we present the results of a probit model. However, the results are similar when using a linear probability model.

[14] For developers with more than one product, product-level covariates are included by averaging across all the developer's products. Multi-product companies often used related products within the same category.

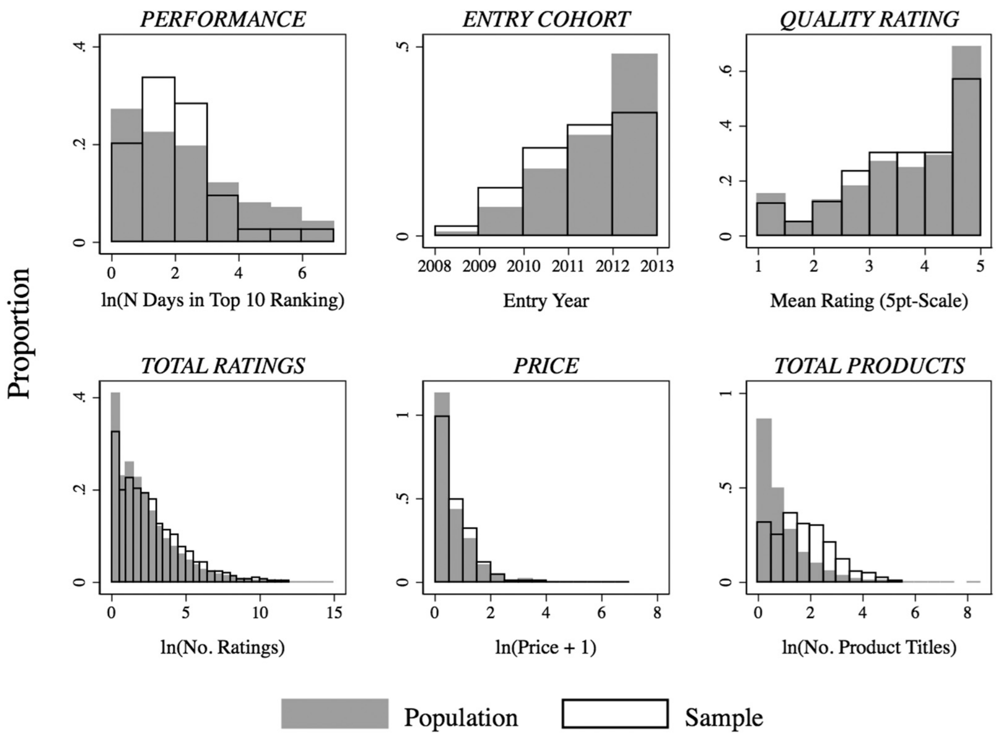

Fig. 1. Distribution of Characteristics for Sample and Population. Note. PRODUCT_SCOPE is a count of the number of products the developer makes available on the App Store. This includes both variants of a given product and entirely different products. This does not include updated (replacement) versions of the same product. The unit of analysis for this comparison is at the product level.

are estimated simultaneously as a bivariate probit regression using maximum likelihood, allowing for correlated error terms. This allows us to control for the other type of IP in each regression.[15] Robust standard errors are reported.

Controls. Indicator variables for product categories (niches) are included, as well as indicators for developer firm size, firm age (market tenure), product rating, and indicators for whether developers were motivated to profit from their products or were creating mobile apps as a hobby.

Strategy to account for omitted variable bias. We use several approaches to account for the possibility of omitted variable bias: (a) Exploit the set of fine-grained developer and product characteristics as control variables; (b) Exploit – as a diagnostic test – the fact that, if omitted variable bias does not play a role, patents should be unrelated to content differentiation (i.e., $\beta_{p,c} \leq 0$), and copyright should be unrelated to design differentiation (i.e., $\beta_{c,d} \leq 0$). Although neither of these may provide a truly causal test, these checks will provide evidence of the robustness of this relationship.

5.1.2. Baseline results

Our analysis is at the developer (firm) level since IP rights are assigned to the firm rather than to the product and are often applied across multiple products. Multiple firms did not reply to particular questions in our survey, and therefore, we did not include them in all the specifications. We also excluded those that were not primarily app developers and attempting to profit from their applications. This includes for example, airlines, food chains, and other companies that offer apps, but whose primary business is not app development.[16] This reduced our sample from 809 respondents to 581 firms.

We present these results in Table 4. In the case of patent use (Column 1), the coefficient for $DESIGN$ (i.e., $\beta_{p,d}$) is positive and statistically significant at 0.29 ($s.e.\ =\ 0.08$); the point estimate for $CONTENT$ (i.e., $\beta_{p,c}$) is statistically indistinguishable from zero at 0.14 ($s.e.\ =\ 0.11$). In the case of registered copyright use (Column 1), the coefficient for $CONTENT$ (i.e., $\beta_{c,c}$) is large and statistically significant at 0.29 ($s.e.\ =\ 0.11$), while the coefficient for $DESIGN$ (i.e., $\beta_{c,d}$) is considerably smaller at 0.18 ($s.e.\ =\ 0.07$) and only marginally significant. Adding the series of dummy controls for firm size, as in model (2), leads this latter coefficient for $DESIGN$ to become statistically indistinguishable from zero at 0.11($s.e.\ =\ 0.07$). These coefficients indicate that a one-point increase in *DESIGN* is associated with a 17% increase in patent use, and a one-point increase in *CONTENT* is associated with a 29% increase in registered copyright use at the centroid of the data (i.e., setting the regressors to mean levels).

Adding other developer or product controls to the model in any combination does not change the results. Column 3 adds other control variables, including a dummy indicating amateur hobbyist developers, e mean product quality rating, and a series of dummies capturing any cases where the developer uses means of capturing revenues other than sales or in-app purchases. The results are also robust to adding patenting as a regressor in the copyright model (and vice versa).

[15] Estimates with linear probability models find similar results. Allowing for correlation in error terms allows us to account for the possibility that, for instance, developers who patent also use copyright, without introducing these terms into the regressions. If these terms are introduced to the regressions, the results remain consistent in sign and statistical significance.

[16] The results did not change when we included the basic covariates with genre controls, performing the regression on the full sample of 809 firms. However, since we show the results including multiple controls, we report results for the sample of firms for which we observed all the variables. Creating weights and applying reweighting to these samples does not change the results of the regressions.

Table 4
Probit Model Results for Patent and Copyright Use.

	(1)	(2)	(3)	(4)
	Bivariate Probit			
Dep. Var.: *PATENT*				
DESIGN	0.29***	0.24**	0.25**	
	(0.08)	(0.09)	(0.09)	
DESIGN = 1				0.06
				(0.23)
DESIGN = 2				0.24
				(0.25)
DESIGN = 3				0.69*
				(0.32)
CONTENT	0.14	0.03	−0.01	
	(0.11)	(0.12)	(0.12)	
CONTENT = 1				−0.05
				(0.22)
CONTENT = 2				0.08
				(0.33)
Dep. Var.: *COPYRIGHT*				
DESIGN	0.18*	0.11	0.14*	
	(0.07)	(0.07)	(0.07)	
DESIGN = 1				0.21
				(0.19)
DESIGN = 2				0.19
				(0.22)
DESIGN = 3				0.44
				(0.28)
CONTENT	0.29**	0.20*	0.19*	
	(0.11)	(0.11)	(0.11)	
CONTENT = 1				0.05
				(0.18)
CONTENT = 2				0.60*
Controls				(0.27)
Controls				
Product Category FEs	Yes	Yes	Yes	Yes
Firm Size FEs		Yes	Yes	Yes
Market Tenure			Yes	Yes
QUALITY RATING				
Revenue Model			Yes	Yes
Motivation			Yes	Yes
N	581	581	581	581
χ^2	29.26	341.64	1311.22	2961.16

Note. Robust standard errors in parentheses (* $p < 0.1$, ** $p < 0.01$, *** $p < 0.001$).

In Column 4, we repeat the analysis but stratify the continuous variable for $DESIGN$ and $CONTENT$ into dummies indicating the level. This is a further test of our hypothesis, which argues that the effect is driven by the degree of differentiation. The probability of patent use increases with each increment of $DESIGN$ differentiation, as does statistical significance ($\beta_{D=1} = 0.06$, $s.e. = 0.23$; $\beta_{D=2} = 0.24$, $s.e. = 0.25$; and $\beta_{D=3} = 0.69$, $s.e. = 0.32$). By comparison, the magnitude of the effect with each increment of $CONTENT$ is much smaller and not statistically significant ($\beta_{C=1} = -0.05$; $s.e. = 0.22$;, $\beta_{C=2} = 0.08$, $s.e. = 0.33$). The probability of copyright use increases with each increment of $CONTENT$ differentiation ($\beta_{C=1} = 0.05$ and $\beta_{C=2} = 0.60$, $s.e. = 0.27$). By comparison, the coefficients with each increment of $DESIGN$ are smaller ($\beta_{D=1} = 0.21$, $s.e. = 0.19$; $\beta_{D=2} = 0.19$; $s.e. = 0.22$; $\beta_{D=3} = 0.44$, $s.e. = 0.28$). However, these coefficients are also increasing and do not provide strong evidence that there is no relationship between patent use and CONTENT. This may be attributed to the fact that certain products may be more likely to be protected in general. We attempt to account for this in the subsequent section.

In Fig. 2, we present marginal effects to aid the interpretation of the effects from Table 4, Column 3. Across a range of specifications, we find evidence that the coefficients $\beta_{p,d}$ and $\beta_{c,c}$ in expressions (1) and (2) are positive. Furthermore, $\beta_{p,c}$ and $\beta_{c,d}$ are each statistically indistinguishable from zero. We should acknowledge that the coefficients for $\beta_{c,d}$ are positive in sign and the magnitude of the coefficient increases for products with more design elements. This finding suggests that the potential omitted variable problem, where firms using design or content differentiation are superior in some way and more likely to use IP rights, does not appear to be the case, at least for fully controlled regressions. Instead, the absence of strong effects for $\beta_{p,c}$ and $\beta_{c,d}$ suggests that we captured much of this difference with our econometric specifications. This suggests that the results support **Hyp 1** and **Hyp 4**.

5.2. IP rights and product revenues

If $PATENTS$ and $COPYRIGHT$ are suitable and provide effective property rights for protecting elements of $DESIGN$ and $CONTENT$, as the above results suggest, then we would expect that the use of these property rights to protect these specific elements of design would lead to higher revenues, consistent with the idea of "effective" property rights (Teece, 1986). In this part of the analysis, we focus on these relationships and perform analysis that is meant to provide a test of **Hyp 2** and **Hyp 5**. However, this is also intended to further corroborate the earlier analysis suggesting that product characteristics and IP use are related.

The unit of observation is the individual product (or app) because product revenues are generated at the level of individual apps, and these apps vary across genres, release dates, prices, and other covariates. We can observe revenues for products that at some point achieve

Table 5
OLS Revenue Regressions for Patents.

Dep. Var.: $REVENUE$	(1)	(2)	(3)	(4)	(5)
$DESIGN$	−0.05	−0.06	−0.07*	−0.05	
	(0.04)	(0.04)	(0.04)	(0.04)	
$PATENT$	0.18	−0.27	0.04	−0.21	0.03
	(0.23)	(0.71)	(0.58)	(0.74)	(0.22)
$PATENT \times DESIGN$		0.23*	0.26*	0.30*	
		(0.13)	(0.14)	(0.14)	
$CONTENT$				0.09	0.08
				(0.06)	(0.05)
$PATENT \times CONTENT$				−0.25	−0.15
				(0.23)	(0.24)
$DESIGN = 1$					0.06
					(0.10)
$DESIGN = 2$					−0.18
					(0.12)
$DESIGN = 3$					−0.18
					(0.12)
$PATENT \times DESIGN == 1$					0.33
					(0.41)
$PATENT \times DESIGN == 2$					1.16*
					(0.46)
$PATENT \times DESIGN == 3$					1.34*
					(0.68)
$CONTENT == 1$					−0.02
					(0.09)
$CONTENT == 2$					0.35*
					(0.21)
$PATENT \times CONTENT == 1$					0.23
					(0.36)
$PATENT \times CONTENT == 2$					−0.94
					(0.66)
$Constant$	2.86***	2.87***	2.43***	2.32***	2.44***
	(0.16)	(0.16)	(0.17)	(0.16)	(0.17)
Controls					
$Product\ Category\ FEs$	Yes	Yes	Yes	Yes	Yes
$Market\ Tenure$	Yes	Yes	Yes	Yes	Yes
$Firm\ Size\ FEs$	Yes	Yes	Yes	Yes	Yes
$Revenue\ Model$			Yes	Yes	Yes
$Other\ IP$			Yes	Yes	Yes
$Product\ Quality\ Rating$			Yes	Yes	Yes
$Motivation$			Yes	Yes	Yes
N	1554	1554	1554	1554	1554
R^2	0.26	0.27	0.24	0.33	0.25
log-likelihood	−2233.42	−2231.90	−2260.48	−2093.06	−2252.92

Note. Clustered (Developer Level) Standard errors in parentheses. (* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$).

a ranking within the top 500.[17] This limits us to a subset of 1554 individual software titles.[18]

We begin with the most direct approach, which is to compare the revenues of firms using either $PATENTS$ or $DESIGN$ to those using both. The interaction of these two variables indicates whether the use of patents to protect elements of digital design yields higher revenues than patents or digital design alone, suggesting that patents are an effective means of protecting innovations. A robustness check then again tests whether the use of patents correlated with products differentiated based on content, suggesting omitted variable bias. If there was no relationship between patents and $CONTENT$, it would indicate that there was no evidence of omitted variable bias. We could then repeat the same analysis for copyrights.

As an illustration of our estimation strategy, we provide an example of estimating equation for patents. We estimate a similar equation for copyright use as well:

$$\begin{aligned} REVENUE_i = {} & \beta_1 PATENTS_i + \beta_2 DESIGN_i \\ & + \beta_3 PATENTS_i \times DESIGN_i \\ & \beta_4 CONTENT + \beta_5 PATENT \times CONTENT_i \\ & + \textbf{OTHER IP}\gamma + \epsilon + \textbf{PRODUCT}_\textbf{i}\delta \\ & + \textbf{DEVELOPER}_\textbf{i}\theta + \epsilon \end{aligned}$$

We estimate the regression using OLS and report the results for patents in Table 5. We report baseline results with only $PATENT$ and $DESIGN$ variables in Column 1, as well as industry categories, market tenure, and firm size controls. We include the interaction in Column 2. The results indicate a positive association with revenues

[17] As described earlier, we imputed revenues from product rankings based on the methodology used by Garg and Telang (2013).

[18] There is an obvious selection based on outcomes issue that arises from using a sample of firms that generate above some threshold in revenues. In this case, the threshold of revenues, particularly in the time period studied around 2013, is close to zero on average. Therefore, we are including the subset of firms generating positive revenues. However, similar to what our earlier discussion of sampling suggests, it is most informative to examine the IP strategies of the firms and products generating positive revenues. However, sampling based on outcomes would bias a simple OLS comparison, which is partly why this is not our preferred approach. The latter analysis based either on simulated counterfactuals or matching provides a like-to-like comparison, which should be unbiased in regards to this issue.

Table 6
OLS Revenue Regressions for Copyright.

Dep. Var.: *REVENUE*	(1)	(2)	(3)	(4)	(5)
CONTENT	0.11*	0.06	0.01	0.01	
	(0.07)	(0.07)	(0.08)	(0.07)	
COPYRIGHT	−0.18*	−0.30**	−0.91*	−1.12**	−0.71*
	(0.10)	(0.12)	(0.38)	(0.35)	(0.43)
COPYRIGHT×CONTENT		0.28*	0.33*	0.31*	
		(0.16)	(0.16)	(0.16)	
DESIGN				−0.08*	
				(0.04)	
COPYRIGHT×DESIGN				0.14	
				(0.13)	
CONTENT = 1					−0.00
					(0.09)
CONTENT = 2					0.14
					(0.20)
COPYRIGHT×CONTENT == 1					0.12
					(0.27)
COPYRIGHT×CONTENT == 2					0.80*
					(0.43)
DESIGN == 1					0.12
					(0.10)
DESIGN == 2					−0.13
					(0.12)
DESIGN == 3					−0.18
					(0.12)
COPYRIGHT×DESIGN == 1					−0.27
					(0.41)
COPYRIGHT×DESIGN == 2					−0.08
					(0.47)
COPYRIGHT×DESIGN == 3					−0.03
					(0.47)
Constant	2.84***	2.85***	2.44***	2.42***	2.41***
	(0.16)	(0.16)	(0.17)	(0.17)	(0.17)
Controls					
Product Category FEs	Yes	Yes	Yes	Yes	Yes
Market Tenure	Yes	Yes	Yes	Yes	Yes
Firm Size FEs	Yes	Yes	Yes	Yes	Yes
Revenue Model			Yes	Yes	Yes
Other IP			Yes	Yes	Yes
Product Quality Rating			Yes	Yes	Yes
Motivation			Yes	Yes	Yes
N	1554	1554	1554	1554	1554
R^2	0.27	0.27	0.24	0.25	0.25
log-likelihood	−2232.13	−2229.59	−2258.75	−2252.20	−2251.31

Note. Clustered (Developer Level) Standard errors in parentheses. (* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$).

($\beta = 0.23$, $s.e. = 0.13$, $p < 0.1$). In Column 3, we introduce additional controls including controls for revenue model (ad-based, in-app, etc.), other IP (copyright, NDAs, EULA), quality rating (both the rating score and interaction of rating score and patents/copyright), and indicators for whether the developer is attempting to generate revenues or is a hobbyist. The results remain consistent ($\beta = 0.26$, $s.e. = 0.14$, $p < 0.1$). In Column 4, we introduce the interaction for $PATENT$ and $CONTENT$, but the results go in the other direction and are not statistically distinguishable from zero ($\beta = -0.25$, $s.e. = 0.23$). The main interaction between $PATENT$ and $DESIGN$ remain consistent ($\beta = 0.30$, $s.e. = 0.14$, $p < 0.1$). This coefficient corresponds to an increase in total revenue of approximately 34% for each increment of design when patent protection is used. As a further check, in Column 5, we split the continuous measure of design differentiation into a series of dummies indicating the degree of $DESIGN$ and $CONTENT$ differentiation. The coefficient for $PATENT$ and the level of $DESIGN$ differentiation is increasing with each increment of design ($\beta_{D=1} = 0.33$, s.e. = 0.41; $\beta_{D=2} = 1.16$, s.e. = 0.46, $p < 0.1$; $\beta_{D=3} = 1.34$, $s.e. = 0.68$, $p < 0.1$). By contrast, the coefficient for $PATENT$ and $CONTENT$ are not increasing with the number of components ($\beta_{C=1} = 0.23$, $s.e. = 0.36$; $\beta_{C=2} = -0.94$, $s.e. = 0.66$). These results provide support for **Hyp 2**.

We report the corresponding results for copyright in Table 6. We report baseline results in Column 1 and introduce key controls in Columns 2 and 3. There is a positive association between the interaction term and revenues ($\beta = 0.33$, $s.e. = 0.16$, $p < 0.1$). In Column 4, we introduce the interaction between $DESIGN$ and $COPYRIGHT$. These results for the interaction between $CONTENT$ and $COPYRIGHT$ are positive and significant ($\beta = 0.31$, s.e. = 0.16, $p < 0.1$), while the relationship between $DESIGN$ and $COPYRIGHT$ is not significant, although the coefficient is positive and smaller ($\beta = 0.14$, s.e. = 0.13). The coefficients indicate an increase of 36% in total revenues for each increment of $CONTENT$ when $COPYRIGHT$ protection is used. In Column 5, we stratify the results again by the level of the variables rather than the continuous variable. We observe an increase in the correlation with revenues with each increment of $CONTENT$ interacted with $COPYRIGHT$ ($\beta_{C=1} = 0.12$, $s.e. = 0.27$; $\beta_{C=2} = 0.80$, $s.e. = 0.43$, $p < 0.1$), but do not observe any increase with each increment of $DESIGN$ that is positive and statistically significant ($\beta_{D=1} = -0.27$, $s.e. = 0.41$; $\beta_{D=2} = -0.08$, $s.e. = 0.47$; $\beta_{D=3} = -0.03$, $s.e. = 0.47$). This suggests that there is a relationship between copyright protection of products that are differentiated on the basis of content and revenues, but we do not observe any relationship between copyright protection and products differentiated based on design and revenues, suggesting

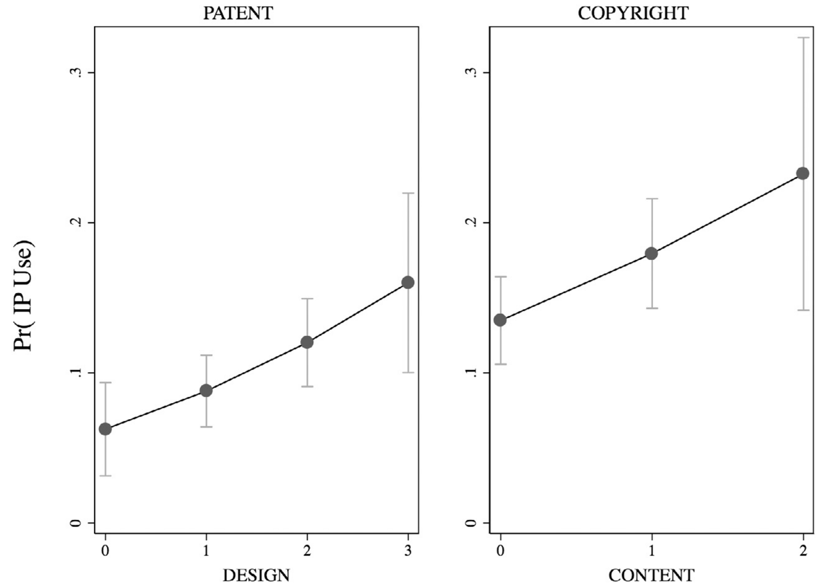

Fig. 2. Marginal Effects for Probit Regressions.

that this is not simply driven by omitted variable bias. These results provide support for **Hyp 4**.

We also test the robustness of our results with a reweighted sample, and the results remain consistent. We tested the robustness of our results using several additional approaches including matching and a simulated counterfactual approach.

5.2.1. *Simulated counterfactual results*

The results in Section 5.2 provide support for **Hyp** 2 and **Hyp 4**. However, one concern is that we do not have a perfect counterfactual for companies with and without patents and differentiation strategies. As an additional robustness check, we use a simulated counterfactual approach.

We develop a framework that attempts to resolve this issue by simulating the underlying counterfactual for each outcome. To illustrate our approach, we describe our approach for the use of $PATENTS$ in relation to the use of the $DESIGN$ variable. We can observe $REVENUE$ and whether firms used $PATENTS$ or $DESIGN$. However, we would ideally want to observe a counterfactual for each of these cases, where we have a product with and without these IP and product characteristics. We can simulate the baseline value, which we would otherwise not be able to observe, by defining the value of the innovation as a latent variable (υ). $PATENTS$ and $DESIGN$ are indicator variables that indicate whether a particular product or firm embodied either of these elements. We can then estimate regression coefficients (β) which indicate whether revenues are greater in the cases where these approaches are used. To illustrate, the basic model we are interested in estimating for patents can be expressed as follows.[19]

$$REVENUE_i = \beta_1 PATENTS_i \upsilon + \beta_2 DESIGN_i \upsilon + \beta_3 PATENTS_i \times DESIGN_i \upsilon + \epsilon$$

Here, i indexes the individual application at the unit of observation. We define υ as a latent variable to be estimated as a function of product and developer characteristics.

$$\upsilon_i = exp(\mathbf{PRODUCT_i}\delta + \mathbf{DEVELOPER_i}\theta + \epsilon)$$

This differs from a conventional regression since these product characteristics are not included as control variables but are instead used to effectively simulate the value of a counterfactual. We show in the Appendix that the results are consistent if we use a basic linear regression with interactions approach.

We define υ as an exponential function, similar to the approach used by Arora et al. (2008), to constrain the baseline revenue generated by a product to be non-negative.[20] To estimate υ, we use the set of developer- and product-level covariates used in the earlier results: product category dummies, firm size, firm age, product quality rating, and a dummy indicating hobbyist developers.[21] Given that our dependent variable here is *REVENUES* (log transformation of total revenues generated by the product), we also include a series of dummy variables to account for other possible revenue sources to ensure that they do not influence the results. These include dummies for advertising business models and auxiliary products and services complementary to app distribution. Their inclusion does not affect the results. Our revenue measure captures revenues from product sales and in-app purchases.

The approach of estimating υ as a latent variable allows us to simulate a counterfactual effectively for each outcome, giving us a way to compare similar companies with and without IP and with different product differentiation strategies. The coefficients, in this case, are not interpreted as regular regression coefficients. Instead, we can interpret them as multipliers that increase the value of the product as a consequence of using a product or IP strategy. Therefore, a coefficient of $\beta = 1$, for example, corresponds to zero effect, while a coefficient statistically greater than zero indicates there is an increase in the level of revenues being generated. The expressions above are both simultaneously estimated using maximum likelihood.

We report the results in Table 7. To summarize, we observe that the interaction between $PATENT$ and $DESIGN$ is associated with a 44% increase in revenues compared to either strategy on its own and no effect for the interaction between $PATENTS$ and $CONTENT$.

[19] This expression can be expanded to include PATENTS and COPYRIGHTS, as well as DESIGN and CONTENT, all within the same expression, with all sixteen possible combinations. The results remain consistent. However, for expositional simplicity, we show the most basic representation with DESIGN and PATENTS.

[20] Replacing the exponential function with a linear function gives similar results.

[21] The results are not particularly sensitive to how R is modeled, as long as product category and developer size are included in the specification. Other combinations of variables included in the model for R provide similar results as those reported here. We report the complete set of variables for consistency with earlier analyses.

Table 7
Revenue Regression Results with Simulated Counterfactual.

Dep. Var.: *REVENUE*	Patent			Copyright		
	(1)	(2)	(3)	(4)	(5)	(6)
Patent Multiplier Estimates (Λ)						
PATENT	1.20** (0.17)	1.02 (0.64)	1.09* (0.18)			
PATENT × *DESIGN*		1.44*** (0.14)				
PATENT × *CONTENT*			1.00 (0.66)			
Copyright Multiplier Estimates (Λ)						
COPYRIGHT				1.00 (0.17)	1.04 (0.83)	1.00 (0.16)
COPYRIGHT × *DESIGN*					01.00 (0.13)	
COPYRIGHT × *CONTENT*						1.23*** (0.12)
DESIGN		1.00 (0.10)			1.00** (0.04)	
CONTENT			1.10*** (0.07)			1.10*** (0.07)
Variables Used to Estimate Counterfactual (R)						
Non Focal IP	Yes	Yes	Yes	Yes	Yes	Yes
Product Category FEs	Yes	Yes	Yes	Yes	Yes	Yes
Firm Size FEs	Yes	Yes	Yes	Yes	Yes	Yes
Market Tenure	Yes	Yes	Yes	Yes	Yes	Yes
Motivation Controls	Yes	Yes	Yes	Yes	Yes	Yes
QUALITY RATING	Yes	Yes	Yes	Yes	Yes	Yes
Revenue Source FEs	Yes	Yes	Yes	Yes	Yes	Yes
N	1554	1554	1554	1554	1554	1554
$log - likelihood$	−2175.32	−2173.65	−2173.33	−2177.80	−2178.39	−2177.49
χ^2	551.01	553.29	561.79	547.83	543.72	545.47

Note. Standard errors in parentheses (* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$).

Similarly, we observe that the combination of $COPYRIGHT$ and $CONTENT$ is associated with revenues that are 23% higher, but we observe no relationship for the relationship between $COPYRIGHT$ and $DESIGN$. We control for other IP strategies in all the models. These results provide further support for **Hyp 2** and **Hyp 4**.

5.2.2. Results with matched sample

As a robustness check, we re-estimate the results using a sample based on coarse exact matching (CEM) (Blackwell et al., 2009). Matching replicates the earlier analysis but on a sample where the firms with and without IP strategies are in similar categories, have a similar market tenure, and the products are released by firms of a similar size. We match separately for $PATENTS$ and $COPYRIGHT$. We performed matching based on product category, firm size, and firm age, as well as IP choice (other than the focal IP choice in the model) and product characteristics ($DESIGN$ or $CONTENT$) not tested directly in the model. Matching reduces the sample only to observations for which a comparable observations exists. The number of observations used to estimate the effect of patents is 154 and 238 for copyright.

The CEM model estimates, including robust standard errors, are reported in Table 8. As Columns 1 and 2 show, the interaction between *PATENTS* and *DESIGN* has a positive and significant relationship with revenues, but *PATENTS* and *DESIGN* are individually not significant. As Column 3 and 4 show, the interaction between *COPYRIGHT* and *CONTENT* has a significant and positive relationship with revenues, but neither *COPYRIGHT* nor *DESIGN* is significantly related to revenues individually. While these results are based on a subset of the data and, therefore, not as general as the earlier results, they confirm the earlier results.

5.3. IP rights and technology trade

As we argue in our hypothesis development, if the use of IP rights is useful and effective in protecting particular types of digital innovations, then they would be critical to fostering a market for technology and encouraging the trade of this technology (Gans and Stern, 2003; Arora et al., 2004). This pattern was also established by the findings of the Carnegie Mellon Survey papers, which found evidence that commercialization strategies and IP were are correlated with technology licensing (Arora and Ceccagnoli, 2006). Through our survey, we captured whether a firm had licensed its products. The variable $LICENSING$ is an indicator variable for whether a developer firm had reported being engaged in the sale or licensing of apps or app technologies. We estimate similar regressions as in Tables 5 and 6. Note that since we are measuring the use of $LICENSING$ at the company level, the regression results are at the firm level, giving us 581 observations. Results for the probit regressions are reported in Table 9.[22]

Columns 1 through 4 include the basic variables. We include the interactions in Column 5 and the full set of controls in Column 6. We observe a strong positive relationship between the interaction of $PATENT \times DESIGN$ and $LICENSING$ ($\beta = 0.88$, $s.e. = 0.26$, $p < 0.001$) but no relationship with $COPYRIGHT$ and $CONTENT$. In Columns 7 and 8, we report the results of the CEM estimator, breaking down estimates in relation to patents and copyright to preserve statistical power and degrees of freedom. These models confirm the significant

[22] Linear models give similar results.

Table 8
Results for Revenue Regressions with Coarse Exact Matching (CEM).

Dep. Var.: *REVENUE*	(1)	(2)	(3)	(4)
PATENT	0.64**	0.12		
	(0.25)	(0.33)		
DESIGN	0.13	0.05		
	(0.11)	(0.13)		
PATENT		0.47**		
× *DESIGN*		(0.19)		
COPYRIGHT			0.03	−0.19
			(0.16)	(0.18)
CONTENT			0.14	−0.10
			(0.18)	(0.18)
COPYRIGHT				0.80**
× *CONTENT*				(0.32)
Matching Variables				
Product Category FEs	Yes	Yes	Yes	Yes
Firm Size FEs	Yes	Yes	Yes	Yes
Market Tenure	Yes	Yes	Yes	Yes
Differentiation Strategy	Yes	Yes	Yes	Yes
Constant	0.73***	0.85***	0.72***	0.80***
	(0.21)	(0.22)	(0.12)	(0.13)
N	154	154	238	238
R^2	0.06	0.09	0.01	0.04
$log - likelihood$	−231.75	−229.82	−352.28	−348.07
F	−3.64	6.70	0.30	2.48
	(0.03)	2.20	(0.74)	(0.06)

Robust standard errors in parentheses (* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$).

Table 9
Results of Probit Regressions for Technology Licensing.

Dep. Var.: *LICENSING*	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Probit						CEM Probit	
PATENT	0.51*				−0.72	−0.76	−0.19	
	(0.21)				(0.42)	(0.51)	(0.35)	
DESIGN		0.15			−0.09	−0.09	0.04	
		(0.09)			(0.10)	(0.10)	(0.08)	
PATENT					0.79***	0.88***	0.56**	
× *DESIGN*					(0.22)	(0.25)	(0.19)	
COPYRIGHT			−0.12		−0.15	−0.13		0.32
			(0.21)		(0.31)	(0.33)		(0.24)
CONTENT				0.33**	0.43**	0.44**		0.40**
				(0.12)	(0.14)	(0.14)		(0.12)
COPYRIGHT					−0.41	−0.40		
× *CONTENT*					(0.29)	(0.32)		−0.14
								(0.23)
Controls:								
Category FEs	Yes	Yes	Yes	Yes	Yes	Yes		
Firm Size FEs	Yes	Yes	Yes	Yes	Yes	Yes		
Market Tenure						Yes		
Product Quality Rating						Yes		
Revenue Model						Yes		
Motivations						Yes		
N	581	581	581	581	581	581	139	168
log-likelihood	−156.95	−158.07	−159.47	−155.60	−145.82	−129.67	−178.10	−184.70
χ^2	69.70	70.34	66.57	86.87	117.58	148.43	27.15	15.37
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)

Note. Robust Standard errors in parentheses (* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$).

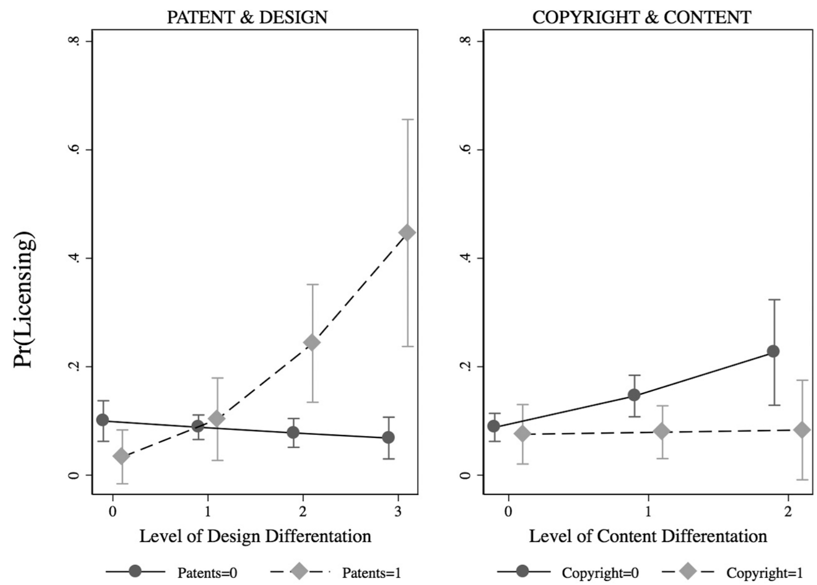

Fig. 3. Marginal Effects for Technology Licensing Regressions.

positive coefficient for patents in conjunction with differentiated design, while the other coefficients remain indistinguishable from zero. This provides support for **Hyp. 3** but not for **Hyp. 6** (Marginal effects reported in Fig. 3).

6. Discussion and conclusions

In this paper, we have studied the effectiveness of patents and copyright as protection strategies in the increasingly important domain of digital innovation. As more and more of the economy has shifted to the digital domain, a growing share of value creation and capture is based on products that are comprised primarily of code, at times combined with other technologies. There are competing narratives regarding the extent to which property rights are effective in this domain. While there is evidence that property rights might be effective (Cockburn and MacGarvie, 2011), there are also questions to what extent property rights are infrequently used (Graham et al., 2009; Miric et al., 2019). Part of the explanation may be attributed to the large number of small firms that exist in digital industries, which may explain the low incidence of property rights. Part of the explanation may also lie with product characteristics, whereby the characteristics of products being protected may shape whether they are protected and by which type of intellectual property. The focus of this paper provides a contrast to earlier studies looking at protection strategies in digital industries, which have focused on evidence at the industry level of the industry (Lerner and Zhu, 2007; Graham et al., 2009).

We argue that product characteristics are likely associated with the choice of whether a product is protected by patents and copyright. Products that are more novel or distinct, which we refer to as design differentiation, are more likely to be protected by patents, and when patents are used to protect these innovations, that is likely to result in higher revenues and a greater likelihood that the products (or a related technology) are subsequently licensed. Products that are more differentiated on the basis of having unique content are more likely to be protected using copyright, and when copyright is used to protect such products, it is likely to result in higher revenues. We find evidence supporting the arguments above but do not find evidence that copyright used to protect products differentiated based on content is associated with greater likelihood of technology licensing.

This paper contributes to the literature that has examined the effectiveness of property rights in protecting innovation, and especially those studies that have looked at digital industries (Hall and Ziedonis, 2001; Cockburn and MacGarvie, 2011; Graham et al., 2009). This paper departs from these earlier studies by looking at the role of product characteristics in shaping the use and effectiveness of property rights in digital settings. As discussed above, much of the existing literature focuses on digital characteristics at the industry level and draws conclusions on how effective are property rights at protecting digital innovations in a general sense (Lerner and Zhu, 2007; Graham et al., 2009). We argue that while these property rights may be more or less effective at protecting digital innovations than those in other industries, there remains important variation within digital industries that explains why some products might be protected in digital industries but not others. We show that this may be a feature of the degree to which products are differentiated and therefore unique or novel. Additionally, it may be a feature of the degree to which these products are comprised not only of code but also combined or “mashed up” with other content.

We also add to the literature on property rights by exploring the effect of registered copyright in digital industries (Cockburn and MacGarvie, 2009, 2011; Wagner and Cockburn, 2010; Graham et al., 2009; Huang et al., 2012). Our results indicate that copyright is an important method for protecting digital innovation but that it protects different types of products than do patents. This suggests that a full picture of digital appropriability should include study the use of both patents and copyright. A further distinguishing characteristic of this study is that we are able to observe patenting and copyright behavior in a way that is not readily captured by matching USPTO databases. Therefore, although this paper does not attempt to establish strong causal relationships it provides insights, in a manner not readily available to alternative empirical approaches into the nuance of how firms use IP rights.

Implications for platform complementers. The use and effectiveness of IP such as patents and copyright are often thought to be an important societal issue and are studied in economics from the perspective of policymakers, but the approaches taken frequently do not directly relate to the question of optimal strategies for firms from the perspective of strategic management. Hence, insights from this paper have considerable implications for understanding competitive strategy and strategic management.

How firms may optimally protect their innovations in digital industries and the extent to which IP is useful or effective are important questions we have sought to answer. The present study provides evidence that IP is both effective and useful for protecting digital innovations, and therefore should play a role in developers' broader "appropriability" or protection strategies, through which they may limit competition and capture revenue. Patents and copyright are potentially important strategies for protecting digital products and allowing complementers to "profit from their digital innovations". However, they are only effective under certain conditions and should not be disregarded in digital industries. Rather this should be considered in relation to product characteristics and other factors.

This also has implications for growing concerns around how much power platform owners have over complementers (Khan, 2016; Kenney and Zysman, 2016). Huang et al. (2012) provide evidence for how property rights could be used by complementers in order to protect themselves from platform owners. While we have theorized how platform complementers could use patents or copyrights to limit competition. The evidence that IP rights are effective can also signal that complementers use these protections in order to limit themselves from being expropriated by platforms.

Implications for platform owners. The question of optimal IP policy is becoming increasingly important for strategic management as innovation shifts to digital platforms (as is the case with this paper) and firms (Apple Inc. in the present case) become policymakers and enforcers. Our paper provides evidence that patents and copyright can be effective at protecting digital innovations, and therefore platform owners (policymakers in this digital space) should consider institutions such as IP enforcement policies in order to foster innovation and allow complementers to capture value from their innovations. However, given that property rights may be effective at protecting digital innovations, it also means that platform owners must also consider that property rights might be used in strategic ways that diminish innovation and the platform ecosystem. Therefore, the results of our paper suggest that property rights may be important in shaping innovation within a platform ecosystem and should be considered as part of a platform strategy and governance.

Acknowledgment

Financial support from London Business School, Copenhagen Business School and Northeastern University.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.respol.2022.104477.

References

Arora, A., Ceccagnoli, M., 2006. Patent protection, complementary assets, and firms' incentives for technology licensing. Manage. Sci. 52 (2), 293–308.
Arora, A., Ceccagnoli, M., Cohen, W.M., 2008. R&D and the patent premium. Int. J. Ind. Organ. 26 (5), 1153–1179.
Arora, A., Fosfuri, A., Gambardella, A., 2004. Markets For Technology: The Economics Of Innovation And Corporate Strategy. MIT Press.
Bessen, J.E., 2003. Patent thickets: Strategic patenting of complex technologies. pp. 1–30.
Bessen, J., Hunt, R.M., 2007. An empirical look at software patents. J. Econ. Manag. Strategy 16 (1), 157–189.
Blackwell, M., Icaus, S., King, G., Porro, G., 2009. Cem: Coarsened exact matching in stata. Stata J. 9 (4), 524–546.
Ceccagnoli, M., 2009. Appropriability, preemption and firm performance. Strateg. Manag. J. 30, 81–98.
Cockburn, I., MacGarvie, M., 2009. Patents, thickets and the financing of early stage firms : Evidence from the software industry. J. Econ. Manag. Strateg. 18 (3), 729–773.
Cockburn, I.M., MacGarvie, M.J., 2011. Entry and patenting in the software industry. Manage. Sci. 57 (5), 915–933.
Cohen, J.E., Lemley, M.A., 2001. Patent scope and innovation in the software industry. Calif. Law Rev. 89 (1), 3–4.
Cohen, W.M., Nelson, R.R., Walsh, J.P., 2000. Protecting their intellectual assets: Appropriability conditions and why US manufacturing firms patent (or not). In: Nber Working Paper Series Protecting. NBER Working Paper Series.
Contigiani, A., Hsu, D.H., Barankay, I., 2018. Trade secrets and innovation: Evidence from the inevitable disclosure doctrine. Strateg. Manag. J. 39 (11), 2921–2942.
Fosfuri, A., Giarratana, M.S., Luzzi, A., 2008. The penguin has entered the building: The commercialization of open source software products. Organ. Sci. 19 (2), 292–305.
Gans, J.S., Stern, S., 2003. The product market and the market for ideas: Commercialization strategies for technology enterpenurs. Res. Policy 32 (2), 333–350.
Garg, R., Telang, R., 2013. Inferring app demand from publicly available data. MIS Q. 37 (4), 1253–1264.
Givon, M., Mahajan, V., Muller, E., 1995. Software piracy: Estimation of lost sales and the impact on software diffusion. J. Market. 59 (1), 29.
Goldfarb, A., Greenstein, S., Tucker, C., 2014. Introduction to "economics of digitization". In: Economics Of Digitization. University of Chicago Press, pp. 13–14.
Graham, S.J.H., Merges, R.P., Samuelson, P., Sichelman, T., 2009. High technology entrepreneurs and the patent system: Results of the 2008 berkeley patent survey. Berkeley Technol. Law J. 24 (4), 1255–1327.
Hall, B.H., MacGarvie, M., 2010. The private value of software patents. Res. Policy 39 (7), 994–1009.
Hall, B.H., Ziedonis, R.H., 2001. The patent paradox revisited: An empirical study of patenting in the U.S. semiconductor industry, 1979–1995. RAND J. Econ. 32 (1), 101.
Heald, P., 2014. How copyright keeps works disappeared. J. Empir. Leg. Stud. 11 (4), 829–866.
Huang, P., Ceccagnoli, M., Forman, C., Wu, D., 2012. Appropriability mechanisms and the platform partnership decision: Evidence from enterprise software. Manage. Sci. 59 (1), 102–121.
Kenney, M., Zysman, J., 2016. The rise of the platform economy. Issues Sci. Technol. 32 (3), 61–69.
Khan, L., 2016. Amazon's antitrust paradox. Yale Law J. 126, 710–805.
Landes, W.M., Posner, R.A., 2003. The Economic Structure Of Intellectual Property Law. Harvard University Press, Cambidge, Mass..
Leiponen, A., Byma, J., 2009. If you cannot block, you better run: Small firms, cooperative innovation, and appropriation strategies. Res. Policy 38 (9), 1478–1488.
Lemley, M.A., 1995. Convergence in the law of software copyright? High Technol. Law J. 10 (1), 1–34.
Lemley, M., Menell, P., Robert Merges, Samuelson, P., 2006. Software and Internet Law. Aspen Publishers, New York, NY..
Lerner, J., Zhu, F., 2007. What is the impact of software patent shifts? Evidence from lotus v. borland. Int. J. Ind. Organ. 25 (3), 511–529.
Lessig, L., 2008. Remix: Making Art And Commerce Thrive In The Hybrid Economy. Penguin.
Luo, H., Mortimer, J.H., 2017. Copyright enforcement: Evidence from two field experiments. J. Econ. Manag. Strateg. 26 (2), 499–528.
Manski, C.F., Lerman, S.R., 1977. The estimation of choice probabilities from choice based samples. Econometrica 45 (8), 1977–1988.
Marco, A.C., Myers, A.F., Graham, S., Agostino, P.D., Apple, K., 2015. The USPTO patent assignment dataset: Descriptions and analysis, Vol. 2015. USPTO Economic Working Paper, (2).
Miric, M., Boudreau, K.J., Jeppesen, L.B., 2019. Protecting their digital assets. Res. Policy 48 (8).
Mortimer, J.H., Nosko, C., Sorensen, A., 2012. Supply responses to digital distribution: Recorded music and live performances. Inf. Econ. Policy 24 (1), 3–14.
Nagaraj, A., 2018. Does copyright affect reuse? Evidence from google books and wikipedia. Manage. Sci. 64 (7), 2973–3468.
Oberholzer, F., Strumpf, K., Journal, S., February, N., Oberholzer-gee, F., 2015. The effect of file sharing on record sales : An empirical analysis, vol. 115. (1), pp. 1–42.
Peitz, M., Waelbroeck, P., 2006. Why the music industry may gain from free downloading - the role of sampling. Int. J. Ind. Organ. 24 (5), 907–913.
Reimers, I., 2016. Can private copyright protection be effective? Evidence from book publishing. J. Law Econ. 59 (2), 411–440.
Samuelson, P., 1993. Fair use for computer programs and other copyrightable works in digital form: The implications of Sony, Galoob and Sega. J. Intellect. Prop. Law 49–118.
Samuelson, P., Scotchmer, S., 2002. The law and economics of reverse engineering. Yale Law J. 111 (7), 1575–1663.
Teece, D.J., 1986. Profiting from technological innovation: Implications for integration, collaboration, licensing and public policy. Res. Policy 15 (6), 285–305.
Varian, H.R., 2005. Copying and copyright. J. Econ. Perspect. 19 (2), 121–138.
Wagner, S., Cockburn, I., 2010. Patents and the survival of Internet-related IPOs. Res. Policy 39 (2), 214–228.
Waldfogel, J., 2012. Copyright protection, technological change, and the quality of new products: Evidence from recorded music since napster. J. Law Econ. 55 (4), 715–740.
Waldfogel, J., Aguiar, L., 2018. Quality predictability and the welfare benefits from new products: Evidence from the digitization of recorded music. J. Political Econ. 126 (2), 492–524.