Technological Progress and Ownership Structure

Heng Geng*

University of Hong Kong

Harald Hau**

University of Geneva and Swiss Finance Institute

Sandy Lai***

University of Hong Kong

January 15, 2016

Abstract

Innovation processes under patent protection generate hold-up problems if complementary patents are owned by different firms. We show that in line with Hart and Moore (1990), shareholder ownership overlap across firms with patent complementarities helps mitigate such hold-up problems and correlates significantly with higher patent investment and more patent success as measured by future citations. The positive innovation effect is strongest for concentrated overlapping ownership and for the cases when the overlapping shareholders are dedicated investors.

JEL Classification: L22, G31, G32

Keywords: Patents, Hold-up Problems, Innovation, Institutional Ownership

*Faculty of Business and Economics, University of Hong Kong, K.K. Leung Building, Pokfulam Road, Hong Kong. E-mail: griffin.gengheng1989@hku.hk.

**University of Geneva, 40 Bd du Pont d'Arve, 1211 Genève 4, Switzerland. Tel.: (++41) 22 379 9581. E-mail: prof@haraldhau.com. Web page: http://www.haraldhau.com.

***Faculty of Business and Economics, University of Hong Kong, K.K. Leung Building, Pokfulam Road, Hong Kong. Tel.: (++852) 3917 4180. E-mail: sandy_lai@hku.hk.

1 Introduction

Although technological progress has been recognized as the main source of long-run economic growth, its relation with corporate ownership structure and property rights in patents is less understood. This paper provides a new empirical perspective on the role of equity ownership structure in attenuating hold-up problems induced by patent protection in the corporate innovation process.

Patent protection provides inventors with exclusive rights to the commercial use of their discoveries. But such discoveries are often part of a larger technological process of interdependent innovations, and the full economic value of a patent might only be unlocked if the innovating firm can simultaneously secure access to many complementary patents. Therefore, patent processes generate a hold-up problem whenever such complementary patents are owned by different firms and ex-ante contracting is incomplete.[1]

Building on the property rights theory of Hart and Moore (1990), we argue that joint equity ownership (i.e. shareholder overlap) of an innovating firm with firms controlling complementary patents can attenuate the hold-up problem and contribute to the patent success of the innovating firm. Two separate channels might promote the internalization of such patent hold-up: First, investors with joint ownership in the downstream and upstream firms could influence management of the downstream firm to internalize future patent rent transfers to the upstream firm and avoid the underinvestment in downstream patents. Second, if such patent rent transfer can only be obtained at an efficiency loss (for example, due to potential patent litigations that retard the commercial adoption of the patent), overlapping investors could contribute to a swift conflict resolution about patent rents, which should also increase ex-ante investment incentives.

[1] Recent economic research has documented a negative impact of recent patent proliferation on R&D investment and follow-on innovation (Heller and Eisenberg, 1998; Bessen and Maskin, 2009; Galasso and Schankerman, 2015) and highlighted abusive patent enforcement by so-called "non-practising entities" (Cohen, Gurun, and Kominers, 2015).

To subject this property-right perspective of patent success to an empirical examination, we combine a large sample of U.S. patent data from the United States Patent and Trademark Office (USPTO) with institutional ownership data from Thomson Reuters for the period 1991–2007. In particular, we track stock ownership not only for the innovating firms, but also for firms owning complementary patents. The complementarities are identified directly from patent filings that explicitly list important precursory patents owned by other firms. By law, each newly filed patent must list precursory (upstream) patents that are technologically related and material to the patentability of the new application.[2] Cited precursory patents thus identify rival property rights which are often complementary assets to the downstream patent. These upstream patents then have to be licensed to the (downstream) innovators for the latter to realize the full value of the new patents (Ziedonis, 2004; Galasso and Schankerman, 2010; Noel and Schankerman, 2013).[3] Our analysis identifies potential patent hold-up based on this list of precursory patents and assumes that the list is exogenously determined by the technology to be patented. Patent examiners frequently add precursory patents to the reference list, suggesting a limited scope in manipulating the reference list of precursory patents by the patent filing firms (Alcácer, Gittelman, and Sampat, 2009).[4]

Our main hypothesis states that joint equity ownership between the downstream innovator and the upstream firms controlling complementary patents attenuates the hold-up problem, increases R&D investment, and contributes to the long-run patent success of the innovating firm. Following

[2]The U.S. patent law requires an invention to be useful, novel, and non-obvious to be patented.

[3]Ziedonis (2004) argues that owners of the (upstream) cited patents are reasonable proxies for the potential licensors of the citing patent. Noel and Schankerman (2013) and Galasso and Schankerman (2010) also suggest that a greater number of upstream assignees can signal a greater number of negotiations and disputes required for the commercialization of the downstream patent.

[4]Patent examiners in USPTO are officially responsible for constructing the list of prior art references. However, inventors also have a "duty of candor" to disclose all material prior art; failure to do so can result in an "inequitable conduct" and the court may render the patent unenforceable. Using data from USPTO for all patents granted over the period 2001–2003, Alcácera, Gittelmanb, and Sampatc (2009) document that examiners insert at least one citation in 92% of patent applications. Overall, examiner citations account for 63% of all citations made by an average patent.

the existing literature, we measure *patent success* by the cumulative citation count of each patent that was filed and granted; this measure can be viewed as the intensive margin of patent production.[5] The extensive margin of patent production is measured by the number of successful patents (i.e., patent applications that are eventually approved by USPTO) a firm files in a given year. Our baseline analysis relates *firm-level patent success* (*ln(1+CITES)*) to *shareholder overlap* (SOL). The former is calculated as the (log of) total number of future citations for all patents a firm files in a given year, whereas the latter measures the size of the overlapping equity ownership held by the group of shareholders that invest in both the patent filing firm and the firms owning upstream complementary patents. Consistent with the *hold-up attenuation hypothesis* of joint equity ownership, we find that SOL emerges as the statistically and economically significant determinant of patent success, and the effect is more pronounced in top three R&D-intensive sectors. SOL is is most strongly related to the extensive margin of patent production: More patents are filed if upstream cited patents are controlled by firms which share many common shareholders with the downstream innovator. Our result holds regardless of whether SOL is measured based on equity ownership overlap in the year just before the patent application or two to four years prior to the application date.

Two related hypotheses are also examined: First, shareholder overlap should represent a more powerful mechanism for hold-up resolution if the overlapping shareholders are dedicated investors characterized by concentrated portfolio positions and a long-term investment horizon. Consistent with this intuition, we find a much stronger effect of shareholder overlap on patent success when such overlap or joint ownership originates in dedicated fund holdings. Second, the concentration of overlapping equity stakes should matter if investors in general face coordination problems. Consequently, if the downstream innovating firm and upstream firms are jointly owned by only a few relatively large shareholders, coordinated action should be easier to organize, and shareholders

[5] See for example Aghion, Van Reenen, and Zingales (2013) for a similar definition of firm-level patent success.

might have stronger incentives to resolve a potential hold-up. In accordance with this prediction, we find that the Herfindahl-Hirschman index of (overlapping) shareholder ownership concentration correlates positively with firm-level patent success beyond the shareholder overlap itself. Thus, coordination and contracting problems are not only an issue with respect to ex-ante bargaining between different patent owners, but also constrain the overlapping shareholders in their effort to overcome the ex-ante bargaining failures over patent rights.

Our firm-level analysis controls for a number of firm characteristics and accounts for firm and time fixed effects in various regression specifications. However, because time-varying unobservable firm-specific factors may still pose an inference problem if they influence both shareholder overlap and patent success, we reproduce our regressions at the patent level while controlling for interacted firm and time fixed effects. These specifications directly compare the success of any two patents (in terms of their future citations) filed by the same firm in the same year. We show that even within the same firm-year, patent success is correlated with the varying degree of hold-up a firm faces with respect to different patents in its patent cohort.

We also examine if a firm's shareholder overlap with its complementary patents is related to higher R&D investment and find an economically strong relationship. Moreover, patent rent internalization by overlapping shareholders creates an agency conflict with non-overlapping shareholders, denoted IO^{NOL}. From the perspective of non-overlapping institutional owners, internalization of rent transfers to upstream patent owners implies R&D overinvestment, which they should oppose. Accordingly, we find that a larger share of non-overlapping institutional ownership in the downstream firm correlates negatively with R&D investment.

To address endogeneity concern, we undertake a placebo test. We replace the shareholder overlap based on true patent citations with a placebo shareholder overlap $SOL\_Placebo$ where we replace any cited upstream firm with a similar firm which is not cited by the downstream firm in a given year. The placebo shareholder overlap features no statistically significant effect on hold-up mitigation and patent success. The placebo measure of shareholder overlap also allows us to address concerns about reverse causality. If investors anticipate patent rents and strategically acquire overlapping ownership shares to benefit from such rents, then patent success may cause shareholder overlap rather than vice versa. Yet event evidence for the evolution of shareholder overlap around the patent filing year shows that shareholder overlap evolves identically for the true SOL and for the $SOL\_Placebo$. This suggests that benefit from patent rents doesn't not seem to have a discernible effect on shareholder overlap.

Lastly, we examine two alternative hypotheses to the hold-up model: Aghion, Van Reenen, and Zingales (2013) highlight the monitoring role of institutional investors and their willingness to support risky R&D investments that typically pay off only in the long run. We therefore include *institutional ownership* IO of the downstream firm in our baseline regression. We find that the variable shows negatively significant correlation with patent success if we include it in our baseline regression. This is consistent with the aforementioned role of (non-overlapping) institutional owners to block revenue internalization by the overlapping shareholders. Second, having tech-savvy shareholders, who invest mainly in innovative firms, might lead to an increase in shareholder overlap and simultaneously constitute a governance advantage for a firm engaged in patent competition. We control for this factor by constructing a firm-level proxy of *shareholder innovation focus* SIF. We find that shareholder innovation focus features a positive relation with a firm's long-run patent success, but that its inclusion does not severely weaken the economic significance of *shareholder overlap* SOL.

To the best of our knowledge, the role of stock market ownership structure in mitigating hold-up problems in patent processes has not been subject to any systematic analysis. Ex-ante contracting about access to auxiliary patents is difficult before the feasibility and commercial potential of a new patent are established. Hold-up expectations should reduce ex-ante investment incentives unless overlapping shareholders internalize such rent extraction through simultaneous ownership in the upstream and downstream firms. Costly patent rent extraction (for which efficiency losses occur) might also be reduced through the power of overlapping shareholders vis-à-vis the upstream firm, justifying higher ex-ante R&D investments. Consequently, shareholder overlap might benefit the minority shareholders of both upstream and downstream firms through a combination of R&D investment in more patents and lower average rent dissipation per patent.

As patent citations are used in firm valuation (Harhoff, Narin, Scherer, and Vopel, 1999), overlapping institutional owners may promote cross-citations between firms in which they also have a joint equity stake. As a robustness check, we filter patent citations to exclude those coming from firms cited by the patent filing firm in any of its current or previous filings. Our results remain robust. We also conduct a number of robustness checks and show that measurement biases with respect to citation counts cannot explain the economically large positive correlations between shareholder overlap and patent success and the level of R&D expenditure.

In the following section, we survey the related literature. Section 3 develops a simple model of patent hold-up in the spirit of Hart and Moore (1990); it develops the main hypotheses and motivates the regression specifications. Section 4 discusses the data. Section 5 presents the main evidence for the role of shareholder overlap on patent success. Section 6 explores the role of shareholder overlap on R&D investment and examines the endogeneity of shareholder overlap. Various robustness checks are undertaken in Section 7, followed by conclusions in Section 8.

2 Related Literature

Notwithstanding its prominence in economic theory, the property rights view of the boundaries of the firm has seen very few empirical applications because of a variety of obstacles. First, noncontractible hold-up problems are often difficult to identify in a complicated business environment. Explicit citation of precursory patents in the patent documents provides a unique identification opportunity. Second, underinvestment at the project level is difficult to measure because a firm can shift investments to other projects for which hold-up problems are less severe. Such an analysis requires a level of disaggregation typically not available for investment data. Third, investments may involve intangibles resources (such as managerial attention) that pose additional measurement problems[6]. For these reasons, we infer the (latent) underinvestment indirectly from diminished project or patent success. Future patent citations provide a sufficiently precise proxy for patent success at the firm and patent level to allow for a comprehensive study of hold-up in the patent process.

Existing studies on patent hold-up problems (e.g., Shapiro, 2001; Ziedonis, 2004; Hall and Ziedonis, 2007) find that licensing agreements are commonly used in practice—yet these might typically concern the ex-post rent allocation. Licensing agreement might involve substantial royalty fees and their negotiation is not a frictionless process. Alternatively, a firm may invent around the patented technology to avoid being held up, but this is not always possible given the cumulative and sequential nature of technological development. There is also evidence that firms seek outright ownership integration via mergers to resolve patent disputes. Such merger cases are often challenged in court and eventually fail for anti-competitive reasons (Creighton and Sher, 2009). Our study suggests that in liquid equity markets, partial ownership integration via ownership overlap may be achieved at lower costs or may already exist if large institutional shareholders happen to hold shares in both firms concerned.

Some studies have argued that common institutional ownership has significant real effects. Azar, Schmalz, and Tecu (2015) find that such shareholder overlap at the industry level can be

[6] One of key assumptions in Hart and Moore (1990) is investment has to be specific to an asset/product such that the realization of the investment cannot be used for other purposes. In the setting of corporate innovation, such applied research is usually done with a product in mind and is asset-specific in this sense.

detrimental to competition. Hansen and Lott (1996) and He and Huang (2014) discuss the coordination role of common shareholders in internalizing conflicts among intraportfolio firms.[7] The extent to which passive institutional shareholders contribute to such intra-industry coordination is still debated (Harford, Jenter, and Li, 2011). Doidge, Dyck, Mahmudi, and Virani (2015) present direct evidence that institutional investors in Canadian equity market coordinate to improve corporate governance collectively. In some cases, activist investors are found to coordinate (otherwise passive) institutional investors in pursuit of common shareholder objectives—making the dichotomy between activist and passive investors less clear-cut (Mullins, 2014; Appel, Gormley, and Keim, 2015).

Patent reform has become a widely debated policy issue. President Obama in his 2014 State of the Union address singled out the patent system as a priority for economic reform. The U.S. administration has pushed USPTO to examine patent requests more rigorously and define their patentable component more narrowly ex ante in order to reduce the reliance on courts to make those determinations ex post.[8] However, Galetovic, Haber, and Levine (2014) argue that there is no evidence that more patent litigations are associated with patent holders stymieing the commercialization of complex technologies or hindering innovation. The evidence in our paper suggests a significant hold-up effect in the corporate innovation process. We argue that shareholder overlap represents an important palliative to hold-up problems with respect to patent investment.

Other empirical work on the determinants of patent success focuses on the role of institutional shareholders. Aghion, Van Reenen, and Zingales (2013) argue that a large share of institutional shareholders is conducive to patent investment as these shareholders tend to pursue a long-run objective. Our evidence shows that it is important to decompose institutional ownership into

[7]Hansen and Lott (1996) document that TIAA-CREF, a pension fund, was actively engaged in resolving litigation between Apple and Microsoft.

[8]A deterioration in the patent environment is sometimes attributed to the growing role of non-practicing entities (NPEs) or "patent trolls," which specialize in the enforcement of patent rights without having a commercial activity of their own. (See, e.g., Cohen, Gurun, and Kominers, 2015.)

the overlapping and non-overlapping components as the latter correlates negatively with long-run patent success. Bena, Ferreira, Matos, and Pires (2015) relate patent success to foreign institutional ownership, but it is unclear whether foreign institutional ownership merely proxies for more *shareholder overlap* in complementary patents as identified in this study. Brav, Jiang, Ma, and Tian (2014) show that hedge fund activism leads to more efficient use of innovative resources and human capital. Our study complements their finding and identifies activist shareholders as an important mechanism to alleviate hold-up problems in innovation. Recent empirical work has also highlighted the complementarity between equity market development and the degree of patent innovation in both the cross-section of countries (Hsu, Tian, and Xu, 2014) and some particular events (Ostinelli, 2014). Insofar as equity market development allows for a better internalization of hold-up problems (through enhanced and adjustable *shareholder overlap*), our study offers a deeper microeconomic interpretation rooted in the theory of the firm for these documented findings.

3 A Model of Patent Investment

3.1 A Simple Benchmark (with No Hold-up Effect)

A risk-neutral firm s can invest into a continuum of patent projects. Each project is represented by the index number p on the interval $[0, \infty)$, where a higher index number corresponds to higher patent development costs. For simplicity, we assume a continuous increasing convex cost function $C(p)$ with $C'(p) > 0$ and $C''(p) > 0$. The present value from commercialization of the patent project, $V_s(p)$, is proportional to the success of the patent proxied by the number of future citation counts $cites_s(p)$. Hence,

$$V_s(p) = \alpha \times cites_s(p), \tag{1}$$

where $cites_s(p)$ is a random variable with the expected value $E[cites_s(p)] = \mu_s$, and $\alpha > 0$ is a constant. The total expected firm value Π_s follows as

$$\Pi_s = \max_{\overline{p}} \int_0^{\overline{p}} [\alpha\mu_s - C(p)]\, dp, \tag{2}$$

where the interval $[0, \overline{p}]$ denotes the range of patent projects the firm pursues. Value maximization implies the first-order condition $\alpha\mu_s = C(\overline{p})$. For a convex cost function $C(p) = cp^b$ $(b > 1)$, we find that

$$\overline{p} = \left(\frac{\alpha\mu_s}{c}\right)^{\frac{1}{b}} \tag{3}$$

characterizes the optimal range of patent production. We summarize the model implications as follows:

Proposition 1: Patent Production without Patent Hold-up

A value maximizing firm optimally invests in the production of patents on the line interval $[0, \overline{p}]$. Given a patent-level expected citation count $E[cites_s(p)] = \mu_s$ that is proportional to each patent's expected value and a convex cost function $C(p) = cp^b$, we find for

(i) the (log) extensive margin of patent production

$$ln[\overline{p}] = \frac{1}{b} ln\frac{\alpha}{c} + \frac{1}{b} ln(\mu_s) \tag{4}$$

(ii) the firm-level (log) citation counts

$$ln[CITES_s] = ln \int_0^{\overline{p}} E[cites_s(p)]dp = \frac{1}{b} ln\frac{\alpha}{c} + \frac{b+1}{b} ln(\mu_s), \tag{5}$$

(iii) the (log) R&D expenditure

$$ln[R\&D\ Exp] = ln\int_0^{\overline{p}} cp^b dp = ln\frac{c}{1+b} + \frac{b+1}{b}ln\frac{\alpha\mu_s}{c}. \quad (6)$$

The firm-level (log) citation count in Eq. (5) is equal to the (log) extensive margin in Eq. (4) plus the (log) intensive margin $lnE[cites_s(p)] = ln(\mu_s)$. Empirically, we can approximate the intensive margin by the average citation count $\overline{cites}_s$ of a firm's patents.

3.2 The Patent Hold-up Effect

Next, we enrich the model setting to account for hold-up problems with respect to the patent value $V_s(p)$. Suppose that commercialization of each patent p requires consent from the owners of upstream patents (p_u, $u = 1, 2, ...N_p$).[9] These upstream patents allow their owners to extract part of the value (through, e.g., license fees) so that the firm's expected patent value decreases. We denote the share of patent value lost to each upstream patent by $L_s(p, p_u)$ and the aggregate value loss by

$$L_s(p) = \sum_{u=1}^{N_p} L_s(p, p_u). \quad (7)$$

The share $L_s(p) \in [0, 1]$ and its component $L_s(p, p_u)$ depend on the "toughness" of bargaining by the owner of the upstream patent p_u. In the ideal case in which the institutional owners of firm s coincide with those of the firms owning (p_u, $u = 1, 2, ...N_p$), no rent extraction should take place so that $L_s(p) = L_s(p, p_u) = 0$. By contrast, maximal rent extraction occurs if there is no overlap in institutional ownership between the downstream innovating firm and the upstream firms. For simplicity, we assume that the ex-ante expectation for value loss is identical for all patents p produced by the same firm, with $E[L_s(p)] = \overline{L}_s$.

Besides the direct value loss due to rent extraction, the hold-up situation might also reduce

[9] Note that p_u does not include any expired patents because they do not pose any threat to the commercialization of the citing patent.

the total value prospect of each individual patent itself. For example, patent litigation may retard the commercial adoption of a patent and jeopardize its long-run success. We assume that the expected number of citations diminishes according to

$$E\left[cites_s(p)\right] = \mu_s[1 - \overline{L}_s]^{\gamma}, \tag{8}$$

where γ denotes the elasticity of the expected patent success (measured by future citation count) to the retained value share, $1 - \overline{L}_s$, with $\gamma \geq 0$. In the special case $\gamma = 0$, patent hold-up does not compromise the overall long-term patent success and instead amounts to only a simple redistribution of future rents. The expected net value from patent p follows as

$$E[V_s(p)] = \alpha[1 - \overline{L}_s]\ E\left[cites_s(p)\right] = \alpha\mu_s[1 - \overline{L}_s]^{1+\gamma}. \tag{9}$$

The optimal investment policy in the hold-up case requires maximization of the expected present value function

$$\max_{\overline{p}_L} \Pi_s = \int_0^{\overline{p}_L} \left[\alpha\mu_s[1 - \overline{L}_s]^{1+\gamma} - C(p)\right] dp, \tag{10}$$

where the optimal patent range $[0, \overline{p}_L]$ has the upper limit

$$\overline{p}_L = \left(\frac{\alpha\mu_s}{c}[1 - \overline{L}_s]^{1+\gamma}\right)^{\frac{1}{b}}. \tag{11}$$

Proposition 2: Patent Production in the Patent Hold-up Case

A firm accounting for an expected value loss $\overline{L}_s$ per patent optimally invests in the production of patents on the line interval $[0, \overline{p}_L]$. Given a patent-level (ex-ante) expected citation count $E[cites_s(p)] = \mu_s[1 - \overline{L}_s]^{\gamma}$, which is proportional to the expected patent value, a convex cost function $C(p) = cp^b$, and an (ex-ante) expected value loss

$\overline{L}_s = E[L_s(p)]$ for each patent due to patent hold-up, we find for

(i) the (log) extensive margin of patent production

$$ln[\overline{p}_L] = \frac{1}{b}ln\frac{\alpha}{c} + \frac{1}{b}ln(\mu_s) + \frac{1+\gamma}{b}ln[1-\overline{L}_s] \tag{12}$$

(ii) the firm-level (log) citation count

$$ln[CITES_s] = \frac{1}{b}ln\frac{\alpha}{c} + \frac{b+1}{b}ln(\mu_s) + \frac{1+\gamma+b\gamma}{b}ln[1-\overline{L}_s], \tag{13}$$

(iii) the (log) R&D expenditure

$$ln[R\&D\ Exp] = ln\frac{c}{1+b} + \frac{b+1}{b}ln\frac{\alpha\mu_s}{c} + (1+\gamma)\frac{b+1}{b}ln[1-\overline{L}_s]. \tag{14}$$

Eqs. (12)–(14) are exactly the same as Eqs. (4)–(6) except for the third term. The third term in Eqs. (12)–(14) features the same (log) loss term $ln[1-\overline{L}_s] < 0$ and captures how the hold-up problem reduces, respectively, the extensive margin, the overall patent success, and R&D expenditure. The hold-up problem also affects the *intensive margin* $E[cites_s(p)]$ of patent production if $\gamma > 0$.

3.3 Patent Hold-up and Shareholder Overlap

The model estimation has to define empirical proxies for the patent-specific hold-up loss $L_s(p)$ and its unconditional expected value $E[L_s(p)] = \overline{L}_s$. We assume that shareholder overlap influences $\overline{L}_s$ through two channels: First, a *transfer internalization channel* implies that management of the downstream firm will only account for the portion of the transfer payments received by the overlapping shareholders but not the portion paid to the upstream firms' other shareholders in its value maximization. Second, a *transfer reduction channel* suggests that if the rent extraction by upstream firms involves frictions that generate costs for overlapping shareholders without a commensurate benefit, overlapping investors would exercise their influence over the upstream firms in favor of swift conflict resolution and therefore reduce the overall patent transfer payments by the downstream firm. Both channels imply that $\overline{L}_s$ should decrease in *shareholder overlap* either because of a smaller proportion of transfer payments that are not accounted for, or because of the reduction of overall monetary transfers, or both.

We can formalize the role of shareholder overlap as follows: Let $O(p)$ be an ownership function that assigns a patent p to a (single) firm owner at time t. The *pairwise (institutional) shareholder overlap* between the downstream patent p and an upstream patent p_u (listed in the patent filings) can be defined as

$$PSOL(p, p_u) = \sum_i \min[w_{i,O(p)}, w_{i,O(p_u)}], \tag{15}$$

where $w_{i,O(p)}$ and $w_{i,O(p_u)}$ are the ownership share (relative to the total institutional ownership of the respective firm) of institutional investor i in, respectively, firms $O(p)$ and $O(p_u)$ at time t. Without loss of clarity, we omit the time index t from all variable expressions in this subsection. We assume the following reduced form for the distributive value loss function associated with the upstream patent p_u cited by patent p:

$$L_s(p, p_u) = \delta w(p_u) \left[1 - PSOL(p, p_u)\right], \tag{16}$$

where weight function $w(p_u)$ measures the importance of the upstream patent p_u relative to all other upstream cited patents of the follow-up patent p. The parameter $\delta \in [0, 1]$ denotes the degree to which separate asset ownership translates into patent revenue sharing; a larger value for δ implies more rent redistribution due to ownership separation. The total redistributed rents to the N_p upstream patent holders aggregate to a redistributive loss for patent p, given by

$$\begin{aligned} L_s(p) &= \sum_{u=1}^{N_p} \delta w(p_u) \left[1 - PSOL(p, p_u)\right] \\ &= \delta \left[1 - \sum_{u=1}^{N_p} w(p_u) PSOL(p, p_u)\right]. \end{aligned} \tag{17}$$

We can define *patent-level shareholder overlap* as

$$sol_p = \sum_{u=1}^{N_p} w(p_u) PSOL(p, p_u). \tag{18}$$

For the N_s patents filed by firm s at year t, we can approximate the average hold-up loss as

$$\begin{aligned} \overline{L}_s &= \sum_{p=1}^{N_s} w(p) L_s(p) \\ &= \delta \left[1 - \sum_{p=1}^{N_s} \sum_{u=1}^{N_p} w(p) w(p_u) PSOL(p, p_u)\right], \end{aligned}$$

where the weight $w(p)$ denotes the relative importance of patent p. The *firm-level shareholder overlap* can be defined as

$$SOL_s = \sum_{p=1}^{N_s} \sum_{u=1}^{N_p} w(p) w(p_u) PSOL(p, p_u), \tag{19}$$

which captures shareholder commonality between firm s and all other firms owning the upstream patents. The hold-up loss term in Proposition 2 can be approximated by

$$ln(1 - \overline{L}_s) \simeq -\overline{L}_s = \delta[SOL_s - 1], \tag{20}$$

and substitution makes the model directly testable. The expression δSOL_s captures the hold-up attenuation through firm-level shareholder overlap relative to a total (non-attenuated) hold-up effect embodied by δ.

A final measurement issue concerns the choice of weights reflecting the relative importance of any patents p and p_u. Empirically, we measure the relative importance by the relative (log) citation count as follows:

$$w(p) = \frac{ln[1 + cites_s(p)]}{\sum_{p=1}^{N_s} ln[1 + cites_s(p)]} \quad \text{and} \quad w(p_u) = \frac{ln[1 + cites(p_u)]}{\sum_{u=1}^{N_p} ln[1 + cites(p_u)]}. \tag{21}$$

In the robustness section (Section 7), we show that an alternative weighting scheme using a (non-parametric) rank measure of future citations $rank(p)$ in Eq. (21) delivers very similar results. The results are also robust to using equal weights.

3.4 Hypotheses

In this subsection, we summarize the main testable hypotheses. Hypotheses H1, H2, H5, and H6 follow directly from the model, whereas Hypotheses H3, H4, and H7 are extensions based on intuitive economic arguments.

H1: Firm Patent Success and Hold-up Attenuation

The patent success of firm s (in terms of future citation $CITES_{s,t}$) for its cohort of patents filed in year t should increase in the *firm-level shareholder overlap* $SOL_{s,t-1}$ between the firm itself and all other firms owning cited upstream patents that pose potential hold-up problems.

In our main empirical analysis, we measure $SOL_{s,t-1}$ based on equity ownership at the end of year $t-1$. In the robustness analysis (Section 7), we further verify our results using ownership measured in years $t-2$ to $t-4$.

H2: Extensive and Intensive Margins of Patent Production

The extensive margin of patent production (proxied by the number of patents $N_{s,t}$ filed by firm s in year t and eventually granted) correlates positively with the *firm-level* *shareholder overlap* $SOL_{s,t-1}$. The intensive margin $\overline{cites}_{s,t}$ (which measures the average citation success of a firm's patents) should also correlate positively with $SOL_{s,t-1}$ if patent hold-up involves not only value redistribution but also (inefficient) value destruction (i.e., $\gamma > 0$).

A straightforward extension of the hold-up hypothesis distinguishes between shareholder types. We conjecture that dedicated shareholders characterized by concentrated equity portfolios and long-term investment horizons should be more willing and/or more capable of exercising their ownership power to resolve patent hold-up than non-dedicated shareholders.

H3: Shareholder Type within Shareholder Overlap

Shareholder overlap should feature a more positive correlation with patent success if the respective overlap is contributed by more dedicated shareholders.

Moreover, a more concentrated ownership among overlapping shareholders might overcome the free-rider problem of costly lobbying and contribute to patent hold-up internalization.

H4: Concentration of Shareholder Overlap

Coordination problems among overlapping shareholders should make more concentrated overlapping equity ownership more effective in internalizing patent hold-up and therefore correlate positively with patent success.

The patent hold-up and its patent-specific attenuation through shareholder overlap should operate not only at the firm level, but also across different patents filed by the same firm in the same year. Accordingly, we can formulate the following within-firm hypothesis:

H5: Patent Success within the Firm

Within a firm's cohort of patents filed in the same year t, those with the largest *patent-level shareholder overlap* (denoted by $sol_{p,t-1}$ and measured with respect to each individual patent p's cohort of upstream cited patents) should feature the largest patent-level success (denoted by $cites_{p,t}$) in terms of the total future citations for the patent.

Hypothesis H5 has the advantage that it can be tested using a rich set of interacted firm-year fixed effects—thereby controlling for time-varying unobservable firm heterogeneity.

Based on Eq. (14), we can summarize the role of shareholder overlap for the ex-ante investment incentives as follows:

H6: Firm R&D Expenditure

A firm's (log) R&D expenditure (denoted by $R\&D\ Exp_{s,t}$) increases in *firm-level shareholder overlap* $SOL_{s,t-1}$.

The interests of the overlapping shareholders for hold-up internalization are generally not shared by other shareholder groups. Their ability to influence the corporate decision process therefore also depends on the strength of other (or non-overlapping) shareholder groups, as stated in Hypothesis H7:

H7: Non-Overlapping Institutional Ownership

Non-overlapping institutional shareholders in the downstream firms should oppose patent rent internalization by the overlapping shareholders because for them it leads to R&D overinvestment. Hence the institutional ownership share of non-overlaping shareholders (IO^{NOL}) should correlate negatively with R&D expenditure conditional on SOL.

4 Data

Our sample combines institutional ownership data with annual patent and citation data for publicly listed firms in the United States. The ownership data are from the Thomson Reuters 13F database. SEC requires all institutional organizations, companies, universities, and so on that exercise discretionary management of investment portfolios over \$100 million in equity assets to report those holdings on a quarterly basis. All common stock positions greater than 10,000 shares or \$200,000 must be reported. Aghion, Van Reenen, and Zingales (2013) document reporting inconsistencies in ownership data prior to 1991, so we only use ownership data from 1991 onwards.

We collect patent and citation information from the latest version of the National Bureau of Economic Research (NBER) Patent Citation database, which includes annual data for patents granted during the period 1976–2006. We further supplement the NBER data with data from Kogan, Papanikolaou, Seru, and Stoffman (2014). The combined data set provides annual patent and citation information for patents granted over the period 1976–2010.[10]

Our measurement of innovation success follows the existing literature (Griliches, Pakes, and Hall, 1988). To distinguish influential innovations from incremental technological discoveries, we use the total number of a patent p's future citations ($cites_{p,t}$) from the patent filing year t to 2010 as our proxy for patent success. A patent will start to receive citations only after it becomes known to others. USPTO currently publishes patent applications 18 months after their filing dates. Such publications generally are not issued for earlier patents (filed before November 29, 2000); therefore, they typically start to receive citations only after they are granted. According to Hall, Jaffe, and Trajtenberg (2001), it takes on average 18 months for a patent's application to be approved and about 95% of successful patent applications are granted within three years of

[10]The data set includes information on the patent number, name of the patent assignee, the number of citations received by a patent, application and grant year of the patent, etc. We thank Professor Noah Stoffman for making the data set available to us.

application. So the lag between patent filing and the first citation ranges from zero to three years in most cases.

At the firm level, we aggregate the count statistic $cites_{p,t}$ to the total number of future patent citations generated by all granted patents filed by firm s in year t, denoted by $CITES_{s,t}$. Self-citations are excluded. Patent and citation counts are set to zero whenever there is no patent or citation information provided in the data. We also examine the extensive margin of patent production $N_{s,t}$, defined as the number of successful patent filings (i.e., patent applications that are eventually granted) by firm s in year t. The corresponding intensive margin is measured by the average cites per patent $\overline{cites}_{s,t}$ (which equals the ratio of $CITES_{s,t}$ to $N_{s,t}$). Because most of these patent-related measures feature highly right skewed distribution, we generally apply a log transformation $ln(1 + X)$ in order to obtain more normally distributed variables for regression analysis.

We adjust carefully for the two truncation problems commonly associated with patent data. First, the patent data set only includes those patents that are eventually granted, so many patent applications filed in 2009 and 2010 and eventually granted beyond 2010 are not included in the data set. To mitigate this patent truncation bias, we use only patent applications up to 2007 in our empirical analysis. Second, patents tend to receive citations over a long period of time, but in our data set we observe the citations only up to 2010. Following Hall, Jaffe, and Trajtenberg (2001, 2005), we correct for the truncation bias in citation counts by estimating the shape of the citation-lag distribution.

The key explanatory variable is the lagged firm-level *shareholder overlap* ($SOL_{s,t-1}$) between the innovating firm and all other firms controlling complementary patents. Calculation of $SOL_{s,t-1}$ follows the definition in Eq. (19) and is based on ownership data at the end of year $t-1$ for patents filed in year t. We measure ownership overlap at year $t-1$ to reduce the scope for reverse causality from patent applications (in year t) to shareholder ownership changes and therefore variations in SOL.

Because expired patents should not create any hold-up problems, we ignore cited patents that are filed 20 years before the application date of the citing patents in constructing SOL.[11] Moreover, we discard firm-year observations when a firm does not have any successful patent application. We track potential hold-up situations only for those cases for which both the patent filing firm and the upstream patent owner are publicly listed firms. Upstream patents owned by private (non-listed) companies do not enter into the SOL measure as we cannot determine any ownership overlap in these cases. Successful patent filings for which none of the upstream patent owners is a publicly listed firm are again discarded from the sample. Generally, downstream firm owners should find it difficult to acquire an overlapping ownership stake in a private firm, thus limiting the scope of the attenuation effect.[12]

The citation count variable as a proxy for patent success has the important advantage that it can be measured not only at the firm level but also at the patent level. Analogously, we can also measure shareholder overlap at both the firm level (SOL) and the patent level (sol). The weighted sum of the *patent-level shareholder overlap* sol amounts to the *firm-level shareholder overlap* SOL, as implied by Eqs. (18) and (19).

We also use a series of control variables, namely a measure of firm size $ln(1 + Assets_{s,t-1})$, the cumulative R&D investment $ln(1 + R\&D_{s,t-1})$, a measure of relative capital intensity $ln(1 + Capital/Labor_{s,t-1})$, and the sales $ln(1 + Sales_{s,t-1})$. For simplicity, $ln(1 + Capital/Labor)$ is abbreviated as $ln(1+K/L)$. To calculate these control variables, we obtain accounting data from

[11] According to USPTO, the 20-year protection period for utility patents starts from the grant date and ends 20 years after the patent application was first filed. The only exception applies to those patents that are filed before June 8, 1995; these patents have a protection period that is the greater of the 20-year term discussed earlier or 17 years from the grant date. (See http://www.uspto.gov/web/offices/pac/mpep/mpep-2700.pdf.)

[12] The exclusion of privately held patents presumably creates a measurement error for SOL unless shareholder overlap with such patent-owning firms is zero. However, we conjecture that such shareholder overlap with privately held companies may indeed be generally negligible.

Compustat and the stock price and shares outstanding data from CRSP.

Our final sample includes all U.S. publicly listed firms that have more than one successful patent application over the sample period 1992–2007. We require each firm to have at least two valid observations because we control for firm fixed effects in our main regression specifications. Our final sample includes $2,697$ firms. We exclude all firm-year observations with missing values for the explanatory variables. The summary statistics are reported in Table 1. The sample features $17,204$ firm-years of patent production involving a total of $582,722$ patents. On average, a firm produces 34 patents per year. The average (median) firm-level shareholder overlap (SOL) is 19.6% (18.4%) with a large standard deviation of 11.6%. The patent-level shareholder overlap (sol) shows an average (median) value of 34.5% (34.5%) with a standard deviation of 16%. The higher mean and standard deviation for the patent-level shareholder overlap is explained by the fact that firms with many patent filings tend to be both larger and feature a higher level of shareholder overlap. The institutional ownership (relative to the total number of shares outstanding of a firm) generally exhibits an upward time trend, from 42.4% in 1992 to 72% in 2007. The shareholder overlap statistics feature a similar upward time trend as well. Time fixed effects are included in all regressions to ensure that the documented shareholder overlap effect does not capture any parallel time trend in patent success. We provide the detailed definitions of all variables in the Appendix.

5 Evidence on Patent Success

5.1 Baseline Specification

Our baseline regression specification follows Eq. (13) and Eq. (20). As some firms in our sample feature patents without any citations, we replace the term $ln[CITES]$ with $ln[1+CITES]$ in our main regression specification as follows:[13]

[13] We note that all results remain qualitatively similar if we restrict the sample to firms with a strictly positive number of citations and use ln[CITES] as the dependent variable.

$$ln[1 + CITES_{s,t}] = \beta_0 + \beta_1 SOL_{s,t-1} + \beta_2 Controls_{s,t-1} + \epsilon_s + \mu_t + \eta_{s,t}, \tag{22}$$

where the coefficient of interest is $\beta_1 = (\frac{1}{b} + \frac{\gamma}{b} + \gamma)\delta \geq 0$. More shareholder overlap with firms holding the upstream patents should boost the downstream innovating firm's patent success as hold-up problems are attenuated.

We estimate Eq. (22) over the period 1992–2007. The citation count $CITES_{s,t}$ for patents filed by firm s in year t includes all future citations up to year 2010. *Shareholder overlap* ($SOL_{s,t-1}$) measures the ownership overlap at the end of year $t-1$ between the innovating firm and all other firms controlling complementary patents. For the choice of control variables, we follow Aghion, Van Reenen, and Zingales (2013) and include the previous R&D investment $ln(1 + R\&D\ Stock_{s,t-1})$, a measure of relative capital intensity $ln(1+K/L_{s,t-1})$, and the firm sales $ln(1+Sales_{s,t-1})$.[14] To better account for firm size measure, we include the firm's total assets $ln(1 + Assets)$.

Table 2 presents the baseline regressions using the dependent variable $ln(1+CITES)$. Columns 1–3 present the results for all firms and Columns 4–6 for firms in three R&D-intensive sectors (pharmaceuticals, computer hardware, and telecommunications equipment).[15] Columns 1 and 4 control for year fixed effects and industry fixed effects based on four-digit SIC codes, whereas Columns 2–3 and 5–6 control for year and firm fixed effects. We report robust t-statistics allowing for two-way clustering at the firm and year (i.e., patent cohort) level.

The baseline regression shows that *shareholder overlap SOL* represents a statistically and economically significant explanatory variable. The point estimate of 1.288 in Column 1 implies that an increase in shareholder overlap by one standard deviation (or 0.116) increases patent success in terms of log firm citation ($ln[1+CITES]$) by 7.4% of its standard deviation of 2.023,

[14] See also Gompers and Metrick (2001) and Hall, Jaffe, and Trajtenberg (2005).

[15] We identify the three R&D-intensive sectors following Bloom, Schankerman, and Van Reenan (2013). Specifically, they are firms in the following sectors: Pharmaceuticals (SIC codes 2834 and 2835), computer hardware (SIC codes 3570, 3571, 3572, 3575, 3576, and 3577), and telecommunications equipment (SIC codes 3661, 3663, and 3669).

suggesting that shareholder overlap has an economically large attenuation effect on patent success. Inclusion of firm fixed effects in Column 2 limits the power of SOL to explain intertemporal variations in patent success within a firm; yet, the point estimate for the SOL coefficient remains significant only at the 10% level. This much weaker statistical significance level is explained by the double inclusion of firm fixed effects and the four firm-level controls, which together absorb much of the variation in patent success. As the control variables in Column 2 may raise endogeneity concerns with respect to the patent process, we also present results with firm fixed effects only in Column 3, which yields indeed a larger regression coefficient of 1.074 for SOL, compared to the coefficient of 0.310 reported in Column 2.

Columns 4–6 repeat these regressions for the three R&D-intensive sectors. As expected, we find a statistically and economically stronger SOL effect in these sectors. Particularly, the regression specifications in Columns 5-6 with firm fixed effects yield statistically and economically much larger more significant point estimates for shareholder overlap. Not surprisingly, shareholder overlap matters most for patent success in those industries which are most patent intensive.

5.2 Intensive versus Extensive Margins

Shareholder overlap may affect intensive and extensive margins differently. Moreover, separate specifications for both reveal different regression parameters. The specification for the intensive margin follows Eqs. (8) and (20), where the expectation term $E[cites(p)]$ is replaced with the empirical sample equivalent $\overline{cites}$. In the log transformation, we use $ln[1 + \overline{cites}]$ rather than $ln[\overline{cites}]$ to include firms whose patents do not receive any citations. Specifically, we estimate the following equation:

$$ln[1 + \overline{cites}_{s,t}] = \theta_0 + \theta_1 SOL_{s,t-1} + \theta_2 Controls_{s,t-1} + \epsilon_s + \mu_t + \eta_{s,t}, \tag{23}$$

where $\theta_1 = \gamma\delta > 0$ captures the positive effect of less patent hold-up due to shareholder overlap. The parameter γ measures the efficiency loss of patent hold-up as opposed to the distributional loss measured by δ. Rejection of $\theta_1 = 0$ in favor of $\theta_1 > 0$ would imply $\gamma > 0$, suggesting that the hold-up problem produces an adverse effect on the average success of the innovating firm's patents, beyond the loss of rent redistribution to the upstream firms.

Table 3, Columns 1–2 summarize the effect of shareholder overlap on the intensive margin. Column 1 excludes firm fixed effects so that both cross- and within-firm variation in shareholder overlap is reflected in the point estimate of 0.432, implying an increase in shareholder overlap by one standard deviation (or 0.116) corresponds to an increase in the average citation count per patent by about 4.5% of its standard deviation. Inclusion of firm fixed effects in Column 2 restricts the identification of the shareholder overlap effect to intertemporal firm variation. Again, the insignificant coefficient for SOL suggests that much of attenuation effect for the intensive margin of patent success coming from the cross-sectional variation is now absorbed by a combination of firm-level controls and the firm fixed effects.

The empirical specification for the extensive margin of patent production follows Eqs. (12) and (20) as

$$ln[1 + N_{s,t}] = \psi_0 + \psi_1 SOL_{s,t-1} + \psi_2 Controls_{s,t-1} + \epsilon_s + \mu_t + \eta_{s,t}, \tag{24}$$

where N corresponds to the $\overline{p}_L$ in our model description in Section 3.2.[16] The model implied coefficient for the hold-up effect is $\psi_1 = (1 + \gamma)\frac{1}{b}\delta > 0$.

The regression results for the extensive margin are presented in Table 3, Columns 3–4. The point estimate of 0.744 in Column 3 suggests a strong economic significance for the shareholder overlap measure; a one-standard-deviation increase in SOL is associated with a 6.4% increase in the number of patents relative to its standard deviation of 1.35. Moreover, the coefficient retains

[16] Similar to Eq. (22) and Eq. (23), we use $ln[1+N]$ rather than $ln[N]$ in Eq. (24), but the results are qualitatively similar under the two alternatives.

its statistical significance in the specification with firm fixed effects in Columns 4.

Overall, the results suggest that shareholder overlap is strongly associated with both more citations for each granted patent (i.e., the intensive margin of patent success) and the number of granted patents (i.e., the extensive margin of patent production). Because both regression coefficients $\theta_1 = \gamma\delta$ and $\psi_1 = (1+\gamma)\frac{1}{b}\delta$ are strictly positive, we conclude that both parameters γ and δ are strictly positive. Overall, the result is consistent with the model presented in Section 3 that patent hold-up not only redistributes rents ($\delta > 0$), but also compromises long-run patent success ($\gamma > 0$).

5.3 Ownership Structure: Two Additional Dimensions

This section explores the relationship between patent success and two different aspects of institutional equity ownership for the innovating firm as stated in Hypotheses H3 and H4 of Section 3.4. First, we examine the role of investor activism among overlapping shareholders (H3); second, we study if the concentration of the overlapping equity stakes matters (H4).

To test Hypothesis H3, we separate institutional investors into (i) dedicated investors and (ii) passive investors based on a combination of portfolio diversification (proxied by Herfindahl-Hirschman Index (HHI)) and portfolio turnover (proxied by churn ratio defined in Gaspar, Massa, and Matos (2005)). At the end of each year, we sort all institutional investors by the HHI (in descending order) and the churn ratio (in ascending order), respectively, and define the combined rank as the sum of HHI rank and churn ratio rank. We label dedicated investors as those in the top half of the combined rank (high concentration and low turnover) and passive investors as those in the bottom half (low concentration and high turnover). The distribution of investor types along the two dimensions of portfolio concentration (i.e. HHI) and portfolio turnover (i.e. churn ratio) is shown in Figure 1, where red and blue points represent dedicated and passive investors, respectively.

Next, we decompose the shareholder overlap of each firm-year according to the two investor types:

$$SOL_{s,t-1} = SOL\_Dedicated_{s,t-1} + SOL\_Passive_{t-1}. \tag{25}$$

Shareholder overlap from dedicated investors (with both concentrated equity stakes and a long investment horizon) is expected to attenuate hold-up problems more effectively than shareholder overlap from the other investor group. The regression result in Table 4, Column 2 confirms this hypothesis. The coefficient for $SOL\_Dedicated$ is at 3.455 more than two times that for SOL in baseline regression (reported in Table 2 and reproduced in Column 1 of Table 4). Shareholder overlap originating in passive investors with diversified portfolios and a short investment horizon shows a much weaker effect on patent success with a coefficient of 0.976. In conclusion, what matters most for patent success is shareholder overlap in complementary patents coming from dedicated shareholders.

Hypothesis H4 concerns the potential coordination problem among the overlapping shareholders. If the downstream innovating firm and the upstream cited firms are jointly owned by a few relatively large shareholders, coordinated action might be easier to organize, and shareholders might have stronger incentives to resolve a potential hold-up. To test this hypothesis, let's consider a downstream patent p filed by firm s in year t and a related upstream patent p_u owned by firm u. Let $i \in I_{p,p_u}$ denote an overlapping investor who at the end of time $t-1$ owns equity shares (relative to total institutional ownership) $w_{i,s}$ and $w_{i,u}$ in firms s and u, respectively. We can define a Herfindahl-Hirschman index (HHI) of shareholder overlap based on overlapping ownership shares $\varpi_i = \min[w_{i,s}, w_{i,u}]$ of all overlapping shareholders $i \in I_{p,p_u}$. Then, we aggregate this shareholder overlap concentration index over all downstream patents p filed by firm s in year t and over their respective upstream patents p_u to obtain a weighted Herfindahl-Hirschman index of ownership concentration of overlapping shareholders, defined as

$$WHHI_{s,t-1} = \sum_{p=1}^{N_s} \sum_{u=1}^{N_p} w(p)w(p_u)HHI_{p,p_u,t-1} \quad , \tag{26}$$

where $w(p)$ and $w(p_u)$ denote (as before) the relative importance weights for patents p and p_u, respectively, and the ownership shares are measured at the end of year $t-1$. $WHHI$ describes how concentrated the overlapping ownership stakes are at the firm level and thus captures the coordination problem among the overlapping investors.

Table 4, Columns 3 includes $WHHI$ as a separate control variable. The estimated coefficient is statistically significant and positive, suggesting that a concentration of joint ownership shares by overlapping shareholders positively correlates with patent success beyond the shareholder overlap SOL itself. The coefficient estimate of 0.591 in Column 3 implies that an increase in the ownership concentration of shareholder overlap by one standard deviation (or 0.174) generates the same effect on patent success as raising SOL by 33.3% relative to its mean $(= [0.174 \times 0.591] \, / \, [1.574 \times 0.196])$. This suggests that coordination problems among dispersed overlapping institutional investors represent an important impediment to the exercise of effective shareholder power.

5.4 Patent-Level Regressions

The firm-level regressions in the previous section control for a variety of observable firm characteristics and firm fixed effects. Yet, time-varying unobservable influences on both patent success and shareholder overlap may still pose a concern for our inference.[17]

In this section, we include the interaction of firm and year fixed effects $\epsilon_{s,t}$. Therefore, identification of the hold-up attenuation effect on patent success relies entirely on the comparison of different patents filed by the same firm in the same year. Different patent filings by the same

[17] For example, media coverage may boost a firm's citation count and simultaneously trigger stock purchases by investors with an investment bias towards technology stocks, thereby increasing the firm's shareholder overlap measure.

firm may list different upstream patents, resulting in patent-specific hold-up and shareholder overlap even within the same firm-year. The patent-specific hold-up attenuation is captured by the patent-level shareholder overlap $sol_{p,t-1}$ in the regression specification

$$ln[1 + cites_{p,t}] = \beta_0 + \beta_1 sol_{p,t-1} + \epsilon_{s,t} + \eta_{p,t}, \tag{27}$$

where $cites_{p,t}$ denotes the future citation count of patent p filed in year t. Similar to the firm-level regressions, all independent variables lag the dependent variable by one year.

The patent-level citation success $cites_{p,t}$ can capture only the intensive margin of patent production, not the extensive margin, unlike the firm-level measure reported in Table 3. In addition, firm-years that feature only one patent application are discarded from the patent-level regression. (Such cases account for about 27% of firm-years in our overall sample.) The patent-level data thus features a strong selection bias toward those firms with many patents—49% of all patent filings are from the 1% most patent-intensive firms (as measured by the total number of patent filings over the sample period) and the other 51% are from the remaining 99% of firms. In Table 5, we present separate regressions for these different groups of patents: Column 1 reports the results for the full sample, Columns 2 and 3, respectively, for the bottom 50% and top 50% of patents from the least and most patent intensive firm-years, respectively, and Column 4 for those patents from the three most R&D-intensive sectors, as defined in Section 5.1.

The *sol* coefficient of 0.225 reported in Column 1 implies that an increase of shareholder overlap *sol* by one standard deviation (0.160) is related to an increase in the patent-level citation count by 2.6% (= 0.225 × 0.160 / 1.360). This modest economic effect mainly represents the hold-up attenuation effect on the intensive margin of the most patent-intensive firms. Column 2–3 show that the estimated coefficient for *sol* is also statistically highly significant at the 1% level. The statistically significant point estimate for *sol* in Column 4 indicates attenuation effect of *sol* is stronger in leading innovative industries. Overall, the result is consistent with Hypothesis H5 that patent success within a firm is also correlated with the patent-specific shareholder overlap *sol*, which differentiates different patents within the same firm-year.

6 Additional Evidence

6.1 R&D Expenditure and Non-overlapping Institutional Ownership

So far our analysis has focused on patent success as the main measure of the hold-up attenuation effect of ownership overlap; yet, the model (Hypothesis H6) also predicts a positive effect of shareholder overlap on R&D investment. To test this prediction, we undertake a linear regression

$$ln[1 + R\&D\ Exp_{s,t}] = \kappa_0 + \kappa_1 SOL_{s,t-1} + \kappa_2 Controls_{s,t-1} + \epsilon_s + \mu_t + \eta_{s,t}, \quad (28)$$

where Eqs. (14) and (20) predict a positive coefficient $\kappa_1 = (1+\gamma)(1+1/b)\delta > 0$. We include the same control variables as in the previous regressions with the exception of $ln(1 + R\&D\ Stock)$, which is excluded because it summarizes past R&D expenditure.

Table 6 reports the regression results. The effect of shareholder overlap is statistically and economically significant in the specifications both without firm fixed effects (Column 1) and with firm fixed effects (Column 2). For example, an increase in shareholder overlap by one standard deviation (or 0.116) in Column 1 increases R&D expenditure by 12% ($= 0.116 \times 1.947/1.884$) of its standard deviation. The hold-up attenuation effect of shareholder overlap on R&D investment is therefore economically important.

In Hypothesis H7 we explore whether the power of hold-up internalization by overlapping shareholders is counterbalanced by the influence of non-overlapping institutional investors, who should oppose what amounts to R&D overinvestment for the standalone investors. Columns 3 and 4 of Table 6 extend the specifications in Eq. 28 to include the non-overlapping institutional ownership IO^{NOL}. Their ownership share is obtained by substracting from the overall institutional ownership IO those institutions that own equity stakes in both downstream and upstream firms (i.e., the institutions that enter positively into the calculation of SOL). In accordance with Hypothesis H7, the coefficient for IO^{NOL} has the predicted negative sign and is statistically highly significant. We conclude that non-overlapping institutional shareholders constrain the hold-up internalization efforts of overlapping shareholders.

6.2 A Placebo Test

Next we propose a placebo test to check whether the relationship between patent success and shareholder overlap is spurious and driven by other unobservable factors. For this purpose we construct a placebo shareholder overlap ($SOL\_Placebo$) measure for which we replace every firm cited as the true upstream patent owner with a placebo firm of similar characteristics. For any firm patent cohort, the placebo firms are matched to the true upstream firms based on the same four-digit SIC industry code and then on the minimal euclidean distance of both firm size ($ln(1 + Assets)$) and firm patent success ($ln(1 + CITES)$)[18]. By construction matched placebo firms aren't cited by any patent filed by the downstream firm in the respective year, which implies that the respective placebo shareholder overlap cannot provide any hold-up relief.

Column 2 of Table 7 confirms this conjecture. Unlike the true shareholder overlap measure, its placebo equivalent ($SOL\_Placebo$) does not feature any statistically significant correlation with patent success. The positive correlation between shareholder overlap and patent success is therefore contingent on picking exactly those firms for the construction of SOL which are cited in the patents of the downstream firm and not an arbitrary groups of similar firms. General unobservable factors influencing both patent success and shareholder overlap are unlikely to account

[18]If an upstream firm cannot find any proper matching in the four-digit SIC industry code in the same year, we then move up to the three-digit SIC industry code. We normalize $ln(1 + Assets)$ and $ln(1 + CITES)$ by their respective industry average in the same year before minimizing the Euclidean distance.

simultaneously for the positive finding for SOL and the negative finding for $SOL\_Placebo$.

6.3 Endogeneity of Shareholder Overlap?

Asset ownership structure might dynamically adjust to patent hold-up and evolve toward the efficient combination of complementary assets. Under private information about future patent hold-up, investors have an incentive to achieve this joint ownership through shareholder overlap—thus internalizing the hold-up problem. We might therefore expect the shareholder overlap between downstream and upstream firms owning complementary patents to increase prior to the public disclosure of patent filings. For each yearly cohort of patents filed between 1991 and 2007, we measure evolution of the average firm-level shareholder overlap relative to the year of the patent filing. For a cohort of downstream patents filed in year t, let $SOL(t,k)$ represent the average shareholder overlap measured based on ownership data at the end of year $t+k$, where $k=-5,-4,...,4,5$. For example, $SOL(t,-3)$ denotes average shareholder overlap between downstream and upstream firms measured based on ownership in year $t-3$ for all patents filed in year t. The average aggregate ownership overlap (measured at lag k) for all patent filing years follows

$$\overline{SOL}(k) = \begin{cases} \frac{1}{17-|k|}\sum_{t=1991+|k|}^{2007} SOL(t,k), & if \ \ -5 \leq k \leq -1 \\ \frac{1}{17-|k|}\sum_{t=1991}^{2007-|k|} SOL(t,k), & if \ \ 0 \leq k \leq 5 \end{cases}$$

and is plotted in Figure 2[19]. As a benchmark, we also plot the evolution of the (average) placebo shareholder overlap defined similar to $\overline{SOL}(k)$ using the corresponding values $\overline{SOL\_Placebo}(k)$.

Around a patent filing year ($k=0$), the average aggregate shareholder overlap $\overline{SOL}(k)$ depicted in red evolves similar to the average aggregate shareholder overlap of the placebo benchmark $\overline{SOL\_Placebo}(k)$ depicted in blue. The vertical line marks two standard deviations around the

[19]We note that the full set of $SOL(t,k)$ cannot be calculated for all years. For example, for patents filed in 1992, we can only calculate $SOL(t,k)$ for $k=-1,0,...,5$. Similarly, for patents filed in 2007, $SOL(t,k)$ can only be calculated for $k=-5,-4,...,0$.

mean value for each of the two measures. The general upward trend of the two curves reflects the increasing institutional ownership in U.S. stock market over the sample period. Importantly, we find no evidence the shareholder overlap $\overline{SOL}(k)$ endogenously reacts in anticipation of patent rents of future patent filing. Instead its evolution mimics that of the placebo shareholder overlap which is by construction devoid of future patent rents. In conclusion, we find no evidence for an endogenous dynamic adjustment of shareholder overlap in the run-up to patent filings. This finding may be the consequence of legal restrictions on insider trading that might limit the scope for stock trading on private information about future hold-up rents.

7 Robustness

7.1 Other Alternative Explanations for the SOL Effect

This section discusses two alternative determinants of patent success and examines whether they can explain the shareholder overlap (SOL) effect we documented in the previous sections. Aghion, Van Reenen, and Zingales (2013) argue that R&D investments have a long time horizon, and a high share of institutional investors allows management to focus on the long-term return on investment. Following their specification, we measure the share of *institutional ownership* (IO) as the relevant proxy for investor patience. As institutional ownership also correlates with our shareholder overlap measure, it could potentially account for the firm-level evidence presented in Sections 5.1–5.3.

The second hypothesis concerns heterogeneous shareholder sophistication about innovation. Some shareholders might bring particular knowledge to the innovation process, allowing for better governance of the innovating firm. In particular, investors can specialize in acquiring stakes in innovative firms with a disproportionate share of patents. Such a *shareholder innovation focus* is directly measurable based on ownership data in a simple three-step procedure. In the first step, we define for each listed firm the *firm innovation focus* (FIF) as the ratio of the future citation count of all patents filed by firm s' in year t to the industry average citation count during the same period. In the second step, we account for all institutional investors i in firm s and calculate their respective *investor innovation focus* (IIF) as the value-weighted average *firm innovation focus* for all stocks s' in their respective investment portfolios except for stock s itself. Formally,

$$IIF_{i,s,t} = \sum_{s' \setminus \{s\}} x_{i,s',t} \; FIF_{s',t} \quad , \tag{29}$$

where $x_{i,s',t}$ represents the value weight of firm s' in the portfolio of institutional investor i at the end of year t. For any individual institutional shareholder primarily investing in innovative firms, the IIF value would be high. In the third step, the *shareholder innovation focus* (SIF) for firm s is defined as the value-weighted average of *investor innovation focus* for all shareholders i in firm s,

$$SIF_{s,t} = \sum_{i} w_{i,s,t} IIF_{i,s,t} \quad , \tag{30}$$

where $w_{i,s,t}$ represents the equity shares held by institutional investor i relative to the aggregate holdings of all institutional investors in firm s at the end of year t. A firm mostly owned by investors with a high IIF should feature a high SIF value. Shareholders' governance competence (proxied by $SIF_{s,t}$) with respect to the innovating firm s should have a positive effect on the firm's patent success.

Table 7 presents the regression results for the two alternative hypotheses. Including both *shareholder overlap* and *institutional ownership* in Column 3, we find a negative relationship between *institutional ownership* and patent success, whereas *shareholder overlap* retains its positive sign and high level of statistical significance. Table 7, Column 4 includes shareholder innovation focus $SIF_{s,t}$ as the third explanatory variable for patent success. Here we find support for the hypothesis that an innovation focus of a firm's shareholders fosters the patent success of the firm. A one standard deviation increase in $SIF_{s,t}$ is associated with an increase in patent success by 4%

of its standard deviation.

To verify that these results are robust, Columns 5–6 report analogous regressions based on the same (smaller) patent sample as that used by Aghion, Van Reenen, and Zingales (2013). We also mimic their specification by using $ln[CITES_{s,t}]$ as the dependent variable and apply the same control variables.[20] Column 5 reproduces their benchmark regression (reported in Table 1, Column 2 of their paper) with the same statistically significant positive coefficient of 0.546 for institutional ownership (IO). When we augment the regression with *shareholder overlap* SOL as an additional explanatory variable in Column 6, *shareholder overlap* remains statistically significant but *institutional ownership* does not.

7.2 Measurement Issues

We subject a variety of measurement choices to a robustness analysis.[21] The first robustness test concerns the truncation nature of patent citations. Our baseline measure of $CITES$ used in the main analysis follows Hall, Jaffe, and Trajtenberg (2001) in scaling the raw future citation count of each patent by a specific factor (see Table 5 of Hall, Jaffe, and Trajtenberg, 2001) that increases in the time span until the terminal year of our sample. The truncated nature of the dependent variable may imply a time-dependent measurement error as the inferred patent success of later cohorts of patents is based on a shorter time span. Lerner, Sorensen, and Stromberg (2011) propose a shorter but more homogeneously truncated citations count over a three-year period immediately after a patent is granted. Following their approach, we define the relative citation

[20] Aghion, Van Reenen, and Zingales (2013) use $ln[CITES_{s,t}]$ as the dependent variable in their benchmark regression in Table 1, Column 2. They include $IO_{s,t}$, $ln(R\&D\ Stock_{s,t})$, $ln(K/L_{s,t})$, and $ln(Sales_{s,t})$ as the regressors. We use the exact same set of variables in our regressions reported in Columns 5–6. Their dataset is available at: https://www.aeaweb.org/articles.php?doi=10.1257/aer.103.1.277.

[21] A detailed documentation on these robustness tests is available as a Web Appendix to this paper on our website, www.haraldhau.com.

count of a patent as

$$cites\_rel_{p,t} = \frac{cites^{3y}_{p,t}}{\frac{1}{N_k}\sum_{p\in k} cites^{3y}_{p,t}} \quad , \tag{31}$$

where patent success is captured by citation count over a three-year period (after the patent is granted) relative to the aggregate citation count of all N_k patents in the same USPTO technology class k. The firm-level measure $CITES\_rel_{s,t}$ follows as the sum of $cites\_rel_{p,t}$ over all N_s successful patents filed by firm s in year t. The importance weights $w(p)$ and $w(p_u)$ in the calculation of shareholder overlap SOL are also based on $cites\_rel_{p,t}$. The modified shareholder overlap variable is denoted by SOL_rel. Notwithstanding these variable modifications, we still find qualitatively similar results for the hold-up attenuation effect of shareholder overlap.

The second robustness test concerns the measurement of shareholder overlap itself. As patent projects might be initiated several years before the application year t, ownership overlap might also be accumulated earlier than year $t-1$ (which is the measurement year for shareholder overlap used in our main analysis). We find that shareholder overlap measured based on equity stakes at the end of years $t-2$, $t-3$, and $t-4$ still produces statistically significant point estimates for SOL, albeit with a smaller economic significance. This finding is consistent with a build-up of joint equity ownership prior to the filing of the downstream patent in year t.

Third, our benchmark measure of firm-level shareholder overlap SOL uses importance weights based on the citation count of patents. As an alternative measure, we replace the log citations count $ln[1+cites_s(p)]$ in Eq. (21) with a rank measure of future citations $rank(p)$ to obtain a new shareholder overlap measure SOL_rank. This alternative measure of shareholder overlap generates very similar regression coefficients for all reported specifications. As another alternative measure, we suspend the importance weights altogether and aggregate all combinations of downstream and upstream patents under equal weights. The resulted shareholder overlap variable SOL_equal is again statistically highly significant.

Fourth, we repeat Columns 1–3 of Table 2 but use $ln(CITES)$ as an alternative dependent variable. We discard all firm-year observations for which there are successful patent applications that do not receive any citation (i.e., $CITES = 0$). The economic significance of SOL is quantitatively similar in this smaller sample; the point estimate suggests that an increase of SOL by one standard deviation increases the standard deviation of citation count $ln(CITES)$ by 7.9%.

Fifth, we also address the concern that the data may feature certain citation biases. For example, firms may be more inclined to quote upstream patents of firms for which the existing shareholder overlap is large. This kind of bias may exist if the quotation of upstream patents is perceived as a value signal and so imply valuation benefits for the firm's own shareholders with joint ownership in those upstream cited firms. Moreover, the quoted upstream firms may in turn be more likely to cite the downstream innovator—amounting to a reciprocal advertisement channel rather than alleviation of a hold-up situation. In order to eliminate such spurious effects from our regression, we exclude all citations that come from (i) firms quoted in patent p or (ii) any firms that the innovating firm has previously quoted since 1976. We find that the statistical significance for the coefficient of *shareholder overlap* at both patent-level and firm-level regressions remain strong for both filters. Finally, we note that any measurement bias in patent citations due to reciprocal patent citations among jointly owned firms cannot easily account for the positive effect of shareholder overlap on firm patent success. Although inflated reciprocal patent citations might help boost (short-term) equity valuations, it is unclear how such potential manipulation would carry over to (costly) R&D investment.

8 Conclusion

This paper provides a property rights perspective on the success of corporate innovation processes. We argue that the success of patents often depends on access to complementary patents not under the direct control of the innovating firm. From a property rights perspective, the "extended boundary" of the innovating firm includes such complementary patents if both the downstream innovator and the upstream firms owning these complementary patents are linked by joint shareholder ownership. This should particularly be the case if such joint shareholder ownership comes from dedicated investors who exercise power over both firms and therefore mitigate the hold-up problem in corporate innovation processes.

Our identification strategy is based on patent documents that directly list related precursory patents, which may have rival patent claims to new products. We define *shareholder overlap* (SOL) as the (importance-weighted) aggregate minimum ownership share that investors own jointly in both the innovating firm and the firms controlling the complementary assets; an innovating firm with a large SOL value can be interpreted as having an extended firm boundary.

We document the role of *shareholder overlap* for patent success at both the firm level and the patent level; it correlates positively with both the intensive and extensive margins of patent production in an economically significant manner. This finding is robust to a variety of control variables and the inclusion of time and firm (or industry) fixed effects. Using interacted firm and time fixed effects, we show that two patents from the same yearly cohort filed by the same firm perform differently depending on their respective (patent-level) shareholder overlap.

Two additional dimensions of ownership structure are also highlighted: First, shareholder overlap coming from more dedicated investors tends to contribute more to the hold-up attenuation—suggesting that the "extended boundary" of the innovating firm also depends on the type of institutional shareholders. Second, the ownership concentration of shareholder overlap matters independently of the overlap level. This could be explained by the existence of coordination and free-rider problems among a large group of overlapping shareholders.

References

[1] Aghion, P., J. Van Reenen, and L. Zingales, 2013, Innovation and institutional ownership, *American Economic Review* 103(1), 277–304.

[2] Alcácer, J., M. Gittelman, and B. Sampat, 2009, Applicant and examiner citations in U.S. patents: An overview and analysis, *Research Policy* 38(2), 415–427.

[3] Appel, I., T. A. Gormley, and D. B. Keim, 2015, Passive investors, not passive owners, Working Paper. Available at: http://ssrn.com/abstract=2475150.

[4] Azar, J., M. C. Schmalz, and, I. Tecu, 2015, Anti-competitive effects of common ownership, Working Paper. Available at: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2427345.

[5] Bena, J., M. A. Ferreira, P. P. Matos, and P. Pires, 2015, Are foreign investors locusts? The long-term effects of foreign institutional ownership, Working Paper. Available at: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2640045.

[6] Bessen, J., and E. Maskin, 2009, Sequential innovation, patents, and imitation, *RAND Journal of Economics* 40(4), 611–635.

[7] Bloom, N., M. Schankerman, and J. Van Reenen, 2013, Identifying technology spillovers and product market rivalry, *Econometrica* 81(4), 1347–1393.

[8] Brav, A., W. Jiang, S. Ma, and X. Tian, 2014, Shareholder power and corporate innovation: Evidence from hedge fund activism, Working Paper. Available at: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2409404.

[9] Cohen, L., U. G. Gurun, and S. D. Kominers, 2015, Patent trolls: Evidence from targeted firms, Working Paper. Available at: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2464303.

[10] Creighton, S. A. , and S. A. Sher, 2009, Resolving patent disputes through merger: A comparison of three potential approaches, *Antitrust Law Journal* 75(3), 657–690.

[11] Doidge, C., A. Dyck, H. Mahmudi, and A. Virani, 2015, Can institutional investors improve corporate governance through collective action? Working Paper. Available at: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2635662.

[12] Galasso, A., and M. Schankerman, 2010, Patent thickets, courts and the market for innovation, *RAND Journal of Economics* 41(3), 472–503.

[13] Galasso, A., and M. Schankerman, 2015, Patents and cumulative innovation: Causal evidence from the courts, *Quarterly Journal of Economics* 130(1), 317–369.

[14] Galetovic, A., S. Haber, and R. Levine, 2014, Patent holdup: Do patent holders holdup innovation? Working Paper. Available at: http://faculty.haas.berkeley.edu/ross_levine/papers/PatentHoldup_7may2014.pdf.

[15] Gaspar, J., M. Massa, and P. Matos, 2005, Shareholder investment horizons and the market for corporate control, *Journal of Financial Economics* 76(1), 135–165.

[16] Gompers, P., and A. Metrick, 2001, Institutional investors and equity prices, *Quarterly Journal of Economics* 116(1), 229–259.

[17] Griliches, Z., A. Pakes, and B. H. Hall, 1988, The value of patents as indicators of inventive activity, NBER Working Paper 2083. Available at: http://www.nber.org/papers/w2083.

[18] Hall, B. H., A. B. Jaffe, and M. Trajtenberg, 2001, The NBER patent citation data file: lessons, insights and methodological tools, NBER Working Paper 8498. Available at: http://www.nber.org/papers/w8498.

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

[20] Hall, B. H., and R. H. Ziedonis, 2007, An empirical analysis of patent litigation in the semiconductor industry, Working Paper. Available at: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.69.5271&rep=rep1&type=pdf.

[21] Hansen, R. G., J. R. Lott, 1996, Externalities and corporate objectives in a world with diversified shareholder/consumers, *Journal of Financial and Quantitative Analysis* 31(1), 43–68.

[22] Harford, J., D. Jenter, and K. Li, 2011, Institutional cross-holdings and their effect on acquisition decisions, *Journal of Financial Economics* 99(1), 27–39.

[23] Harhoff, D., F. Narin, F. M. Scherer, and K. Vopel, 1999, Citation frequency and the value of patented inventions, *Review of Economics and Statistics* 81(3), 511–515.

[24] Hart, O., and J. Moore, 1990, Property rights and the nature of the firm, *Journal of Political Economy* 98(6), 1119–1158.

[25] He, J., and J. Huang, 2014, Product market competition in a world of cross ownership: Evidence from institutional blockholdings, Working Paper, Available at: http://ssrn.com/abstract=2380426.

[26] Heller M. A., and R. S. Eisenberg, 1998, Can patents deter innovation? The anticommons in biomedical research, *Science* 280(5364), 698–701.

[27] Hsu, P., X. Tian, and Y. Xu, 2014, Financial development and innovation: Cross-country evidence, *Journal of Financial Economics* 112(1), 116–135.

[28] Kogan, L., D. Papanikolaou, A. Seru, and N. Stoffman, 2014, Technological innovation, resource allocation and growth, Working Paper. Available at: http://ssrn.com/abstract=2193068.

[29] Lerner, J., M. Sorensen, and P. Stromberg, 2011, Private equity and long-run investment: The case of innovation, *Journal of Finance* 66(2), 445–477.

[30] Mullins, W., 2014, The governance impact of index fund: Evidence from regression discontinuity, Working Paper. Available at: http://www.rhsmith.umd.edu/files/Documents/Departments/Finance/spring2014/mullins.pdf.

[31] Noel, M., and M. Schankerman, 2013, Strategic patenting and software innovation, *Journal of Industrial Economics* 61(3), 481–520.

[32] Ostinelli, D., 2014, The big innovation bang, Working Paper, University of Zurich.

[33] Shapiro, C., 2001, Navigating the patent thicket: Cross-licenses, patent pools, and standard-setting., A. Jaffe, J. Lerner, S. Stern, eds. *Innovation Policy and the Economy.* NBER, Cambridge, MA.

[34] 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

Reported are the summary statistics for all regression variables in the sample period 1992–2007. Dependent firm-level variables are (i) $CITES_{s,t}$ as number of future citations received by the cohort of patents successfully filed by firm s in year t, (ii) $N_{s,t}$ as the number of successfully filed patents, (iii) $\overline{cites}_{s,t}$ as the average future citations per patent for the cohort of patents filed by firm s in year t, and (iv) $R\&D\ Exp_{s,t}$ as R&D expenditure for firm s in year t. At the patent level, (v) $cites_{p,t}$ denotes the total number of future citations (exclusive of self-citations) received by patent p successfully filed in year t. The explanatory variables $SOL_{s,t-1}$ and $sol_{p,t-1}$, refer to the *shareholder overlap* for firm s or a patent p, respectively. We separate shareholder overlap into SOL contributed by dedicated investors ($SOL\_Dedicated$) and by passive investors ($SOL\_Passive$). $SOL\_Placebo$ denotes a placebo measure for shareholder overlap. $IO_{s,t-1}$ is the aggregate *institutional ownership* of firm s as of the end of year $t-1$. $IO^{NOL}_{s,t-1}$ is the institutional ownership excluding institutions that enter into the calculation of SOL. The *shareholder innovation focus* $SIF_{s,t-1}$ is defined as the investment bias of a firm's shareholders toward firms with a large share of patents. $WHHI_{s,t-1}$ represents the weighted HHI of shareholder overlap concentration for firm s in year $t-1$. The control variables include the (log of) lagged total assets, $ln(1+Assets_{s,t-1})$; lagged cumulative R&D investment, $ln(1+R\&D\ Stock_{s,t-1})$; lagged capital to labor ratio, $ln(1+K/L_{s,t-1})$; and lagged sales, $ln(1+Sales_{s,t-1})$. The variable definitions are described in detail in the Appendix.

	Obs.	Mean	Median	STD	Skewness	Min.	P10	P90	Max.
Dependent Variables (measured in year t)									
$ln(1+CITES)$	17,204	4.154	4.138	2.023	0.056	0.000	1.629	6.736	11.640
$ln(1+N)$	17,204	2.087	1.792	1.350	1.259	0.693	0.693	4.043	8.395
$ln(1+\overline{cites})$	17,204	2.444	2.504	1.112	−0.181	0.000	1.020	3.802	6.278
$ln(1+R\&D\ Exp)$	16,647	2.812	2.778	1.884	0.412	0.000	0.000	5.260	9.408
$ln(1+cites)$	464,486	1.966	2.031	1.360	0.077	0.000	0.000	3.719	6.969
Independent Variables (measured in year $t-1$)									
SOL	17,204	0.196	0.184	0.116	0.527	0.000	0.054	0.354	0.727
$SOL\_Dedicated$	17,204	0.032	0.025	0.029	1.209	0.000	0.000	0.072	0.212
$SOL\_Passive$	17,204	0.157	0.150	0.093	0.466	0.000	0.040	0.282	0.666
$SOL\_Placebo$	17,204	0.148	0.144	0.087	0.460	0.000	0.038	0.262	0.572
sol	464,486	0.345	0.345	0.160	0.065	0.000	0.125	0.555	0.851
SIF	17,204	0.201	0.200	0.054	2.464	0.000	0.140	0.264	1.934
IO	17,204	0.487	0.510	0.265	−0.109	0.000	0.104	0.826	1.000
$WHHI$	17,204	0.198	0.133	0.174	2.045	0.000	0.060	0.430	1.000
IO^{NOL}	17,204	0.069	0.028	0.102	2.632	0.000	0.000	0.195	0.894
Controls (measured in year $t-1$)									
$ln(1+Assets)$	17,204	5.886	5.706	2.168	0.384	0.564	3.243	8.895	13.929
$ln(1+R\&D\ Stock)$	17,204	3.932	4.054	2.205	0.011	0.000	0.026	6.681	10.714
$ln(1+K/L)$	17,204	4.433	4.349	0.885	0.570	0.000	3.428	5.534	8.750
$ln(1+Sales)$	17,204	5.544	5.546	2.462	−0.059	0.000	2.311	8.759	12.722

Table 2: Baseline Regressions

Reported are firm-level OLS regressions of patent success (measured as the (log) future citation count, $ln(1 + CITES_{s,t})$ for all patents filed by firm s in year t) on the lagged *shareholder overlap*, $SOL_{s,t-1}$, for the sample period $1992 - 2007$. Column 1–3 report full sample results whereas Column 4–6 report subsample results based on the top three R&D-intensive industries. Shareholder overlap measures the average shareholder ownership overlap between the innovating firm and other firms owning the precursory complementary patents. The control variables include the (log of) lagged total assets, $ln(1 + Assets_{s,t-1})$; lagged cumulative R&D investment, $ln(1 + R\&D\ Stock_{s,t-1})$; lagged capital to labor ratio, $ln(1 + K/L_{s,t-1})$; and lagged sales, $ln(1 + Sales_{s,t-1})$. Industry fixed effects are based on four-digit SIC codes. All regressions report robust t-statistics clustered at firm and year levels in brackets. We denote by *, **, and *** the statistical significance at the 10%, 5%, and 1% level, respectively. The variable definitions are described in more detail in the Appendix.

Dependent Variables:	$ln(1 + CITES)$					
	Full Sample			Top 3 R&D-Intensive Industries		
	(1)	(2)	(3)	(4)	(5)	(6)
SOL	1.288***	0.310*	1.073***	1.360***	0.939***	2.042***
	[7.31]	[1.75]	[6.28]	[4.11]	[2.86]	[6.59]
Controls:						
$ln(1 + Assets)$	0.279***	0.246***		0.396***	0.221***	
	[12.62]	[8.80]		[10.76]	[4.92]	
$ln(1 + R\&D\ Stock)$	0.432***	0.140***		0.405***	0.264***	
	[34.16]	[4.60]		[14.33]	[4.35]	
$ln(1 + K/L)$	−0.039*	−0.139***		−0.002	−0.078	
	[−1.91]	[−4.27]		[−0.06]	[−1.37]	
$ln(1 + Sales)$	−0.078***	−0.060**		−0.097***	−0.039	
	[−4.61]	[−2.38]		[−3.92]	[−1.04]	
Year FE	YES	YES	YES	YES	YES	YES
Industry FE	YES	NO	NO	YES	NO	NO
Firm FE	NO	YES	YES	NO	YES	YES
Obs.	17,204	17,204	17,204	5,470	5,470	5,470
Adj. R^2	0.517	0.728	0.724	0.555	0.750	0.745

Table 3: Intensive versus Extensive Margin

Reported are OLS regressions for (i) the intensive margin, $ln(1+\overline{cites}_{s,t})$, and (ii) the extensive margin, $ln(1+N_{s,t})$, of patent production on the lagged *shareholder overlap*, $SOL_{s,t-1}$, for the sample period $1992-2007$. We denote by $N_{s,t}$ the number of successful patents filed by firm s in year t, and by $\overline{cites}_{s,t}$ the average future citations per patent for the cohort of patents successfully filed by firm s in year t. Shareholder overlap measures the average shareholder ownership overlap between the innovating firm and other firms owning the precursory complementary patents. The control variables include the (log of) lagged total assets, $ln(1+Assets_{s,t-1})$; lagged cumulative R&D investment, $ln(1+R\&D\ Stock_{s,t-1})$; lagged capital to labor ratio, $ln(1+K/L_{s,t-1})$; and lagged sales, $ln(1+Sales_{s,t-1})$. Industry fixed effects is based on four-digit SIC codes. All regressions report robust t-statistics clustered at firm and year levels in brackets. We denote by *, **, and *** statistical significance at the 10%, 5%, and 1% level, respectively. The variable definitions are described in more detail in the appendix.

Dependent Variables:	$ln(1+\overline{cites})$		$ln(1+N)$	
	(1)	(2)	(3)	(4)
SOL	0.432***	−0.112	0.744***	0.389***
	[4.18]	[−0.96]	[7.20]	[4.27]
Controls:				
$ln(1+Assets)$	0.030**	0.021	0.212***	0.192***
	[2.17]	[1.13]	[17.15]	[13.30]
$ln(1+R\&D\ Stock)$	0.027***	−0.036*	0.367***	0.166***
	[3.56]	[−1.93]	[47.35]	[9.39]
$ln(1+K/L)$	−0.041***	−0.103***	0.003	−0.026
	[−3.20]	[−4.73]	[0.24]	[−1.56]
$ln(1+Sales)$	−0.043***	−0.058***	−0.023**	0.001
	[−4.09]	[−3.37]	[−2.56]	[0.08]
Year FE	YES	YES	YES	YES
Industry FE	YES	NO	YES	NO
Firm FE	NO	YES	NO	YES
Obs	17,204	17,204	17,204	17,204
Adj. R^2	0.437	0.610	0.616	0.834

Table 4: Hold-Up Resolution by Investor Types and Concentration of Shareholder Overlap

The baseline regression in Table 2 is repeated for a split of *shareholder overlap* ($SOL_{s,t-1}$) based on the concentration and turnover of institutional investors' fund holding. The investors are evenly separated into dedicated investors (high concentration and low turnover) and passive investors (low concentration and high turnover). The regression in Column 3 expands the baseline regression by a direct measure of Herfindahl-Hirschman index of shareholder overlap, $WHHI_{s,t-1}$. The control variables include the (log of) lagged total assets, $ln(1 + Assets_{s,t-1})$; lagged cumulative R&D investment, $ln(1 + R\&D\ Stock_{s,t-1})$; lagged capital to labor ratio, $ln(1 + K/L_{s,t-1})$; and lagged sales, $ln(1 + Sales_{s,t-1})$. All regressions report robust t-statistics clustered at firm and year levels in brackets. The last row of the table reports p-values for the null hypothesis that the estimated regression coefficients are the same for $SOL\_Dedicated$ and $SOL\_Passive$. We denote by *, **, and *** the statistical significance at the 10%, 5%, and 1% level, respectively. The variable definitions are described in more detail in the Appendix.

Dependent Var.:	$ln(1 + CITES)$		
	(1)	(2)	(3)
SOL	1.288***		1.574***
	[7.31]		[8.64]
$SOL\_Dedicated$		3.455***	
		[5.32]	
$SOL\_Passive$		0.976***	
		[4.58]	
$WHHI$			0.591***
			[6.95]
Controls:			
$ln(1 + Assets)$	0.279***	0.277***	0.299***
	[12.62]	[12.55]	[13.39]
$ln(1 + R\&D\ Stock)$	0.432***	0.429***	0.432***
	[34.16]	[33.78]	[34.17]
$ln(1 + K/L)$	−0.039*	−0.038*	−0.038*
	[−1.91]	[−1.87]	[−1.88]
$ln(1 + Sales)$	−0.078***	−0.078***	−0.080***
	[−4.61]	[−4.61]	[−4.73]
Year FE	YES	YES	YES
Industry FE	YES	YES	YES
Firm FE	NO	NO	NO
Obs	17,204	17,204	17,204
Adj. R^2	0.517	0.517	0.518
$p - value$		0.001	

Table 5: Patent-Level Regressions

This table presents the correlation between patent success measured at the patent level and the lagged *shareholder overlap* for the sample period $1992 - 2007$. Patent success is proxied by $ln(1 + cites_{p,t})$ as the (log) future citation count received by a patent p filed in year t. Columns 1 reports the full sample result. Columns 2–3 feature the subsamples of 50% patents attributable to top and bottom innovative firm-year observations. Column 4 reports subsample results of the top three industries with highest R&D expenditures. Technology field fixed effect is based on six broad technological categories classified by Hall, Jaffe, and Trajtenberg (2001). The sample excludes firm-years that feature only one patent application as firm fixed effects are in place in all regression specifications. All regressions report robust t-statistics clustered at firm and year levels in brackets. We denote by *, **, and *** the statistical significance at the 10%, 5%, and 1% level, respectively. The variable definitions are described in more detail in the Appendix.

Dependent Variables:	$ln(1 + cites)$			
		Sort by Firm Patent Intensity		Top 3 R&D-intensive
	Full Sample	Bottom 50%	Top 50%	industries
	(1)	(2)	(3)	(4)
sol	0.225***	0.111***	0.277***	0.312***
	[9.44]	[3.57]	[8.77]	[8.33]
Tech. FE	YES	YES	YES	YES
Year × Firm FE	YES	YES	YES	YES
Obs	464,486	232,839	231,647	189,807
Adj. R^2	0.329	0.358	0.298	0.315

Table 6: R&D Expenditure and Shareholder Overlap

Reported are OLS regressions of R&D expenditure for the sample period 1992 – 2007. Dependent variable (log) R&D expenditure is measured for every firm-year (s,t). $SOL_{s,t-1}$ represents firm-level shareholder ownership overlap with all cited firms in the successful patent applications of firm s in year $t-1$. $IO^{NOL}_{s,t-1}$ denotes the (non-overlapping) institutional ownership share in the downstream firm. The control variables include the (log of) lagged total assets, $ln(1+Assets_{s,t-1})$; lagged capital to labor ratio, $ln(1+K/L_{s,t-1})$; and lagged sales, $ln(1+Sales_{s,t-1})$. Industry fixed effects is based on four-digit SIC codes. All regressions report robust t-statistics clustered at firm and year levels in brackets. We denote by *, **, and *** the statistical significance at the 10%, 5%, and 1% level, respectively. The variable definitions are described in more detail in the Appendix

Dependent Variable:	$ln(1 + R\&D\ Exp)$			
	(1)	(2)	(3)	(4)
SOL	1.947***	0.568***	1.498***	0.510***
	[14.47]	[6.65]	[11.02]	[5.89]
IO^{NOL}			−1.247***	−0.161***
			[−13.81]	[−2.71]
Controls:				
$ln(1 + Assets)$	0.742***	0.447***	0.750***	0.449***
	[49.92]	[32.43]	[50.57]	[32.52]
$ln(1 + K/L)$	0.056***	−0.063***	0.052***	−0.063***
	[3.49]	[−4.09]	[3.25]	[−4.08]
$ln(1 + Sales)$	−0.123***	0.065***	−0.120***	0.065***
	[−10.50]	[5.32]	[−10.28]	[5.35]
Year FE	YES	YES	YES	YES
Industry FE	YES	NO	YES	NO
Firm FE	NO	YES	NO	YES
Obs	16,647	16,647	16,647	16,647
Adj. R^2	0.720	0.940	0.724	0.940

Table 7: Alternative Explanatory Variables

We compare four potential determinants of innovation success, namely (i) *shareholder overlap* ($SOL_{s,t-1}$) between an innovating firm and upstream firms owning complementary patents as a proxy for attenuation of a patent hold-up problem; (ii) A placebo measure for shareholder overlap ($SOL\_Placebo_{s,t-1}$) for which we replace every firm cited as the true upstream patent owner with a placebo firm of similar firm characteristics; (iii) *institutional ownership* ($IO_{s,t-1}$) as advocated by Aghion, Van Reenen and Zingales (2013) as a proxy for investor patience; and (iv) *shareholder innovation focus* ($SIF_{s,t-1}$) as a proxy for a firm's shareholders focus on research intensive portfolio investments. Columns 1–4 use the full sample period 1992 − 2007, and Columns 5–6 use the sample of Aghion, Van Reenen and Zingales (2013), which spans the shorter period from 1991 to 1999. The dependent variable $ln(1+CITES_{s,t})$ is the (log) number of total future citations received by the cohort of patents successfully filed by firm s in year t. The first four regressions adopt the same dependent and control variables as in the previous tables. The last two regressions are based on dataset provided by Aghion, Van Reenen and Zingales (2013). Robust t-statistics are reported in brackets. We denote by *, **, and *** statistical significance at the 10%, 5%, and 1% level, respectively. The variable definitions are described in more detail in the Appendix.

Dependent Variables:	$ln(1+CITES)$ Full Sample				$ln(CITES)$ ARZ Sample	
	(1)	(2)	(3)	(4)	(5)	(6)
SOL	1.288***		1.418***	1.252***		2.517***
	[7.31]		[8.02]	[6.93]		[3.93]
$SOL\_Placebo$		−0.275				
		[−1.42]				
IO			−0.481***	−0.472***	0.546***	0.230
			[−8.01]	[−7.86]	[2.94]	[1.04]
SIF				1.480***		
				[4.55]		
Controls:						
$ln(1+Assets)$	0.279***	0.335***	0.309***	0.293***		
	[12.62]	[15.31]	[13.74]	[12.90]		
$ln(1+R\&D\ Stock)$	0.432***	0.439***	0.427***	0.426***	0.337***	0.344***
	[34.16]	[34.80]	[33.71]	[33.64]	[8.43]	[6.58]
$ln(1+K/L)$	−0.039*	−0.040*	−0.041**	−0.041**	0.261***	0.317***
	[−1.91]	[−1.93]	[−2.02]	[−2.01]	[3.09]	[3.29]
$ln(1+Sales)$	−0.078***	−0.075***	−0.071***	−0.066***	0.310***	0.227***
	[−4.61]	[−4.41]	[−4.20]	[−3.88]	[6.94]	[3.86]
Year FE	YES	YES	YES	YES	YES	YES
Industry FE	YES	YES	YES	YES	YES	YES
Firm FE	NO	NO	NO	NO	NO	NO
Obs	17,204	17,204	17,204	17,204	4,025	3,202
Adj. R^2	0.517	0.515	0.519	0.520	0.611	0.626

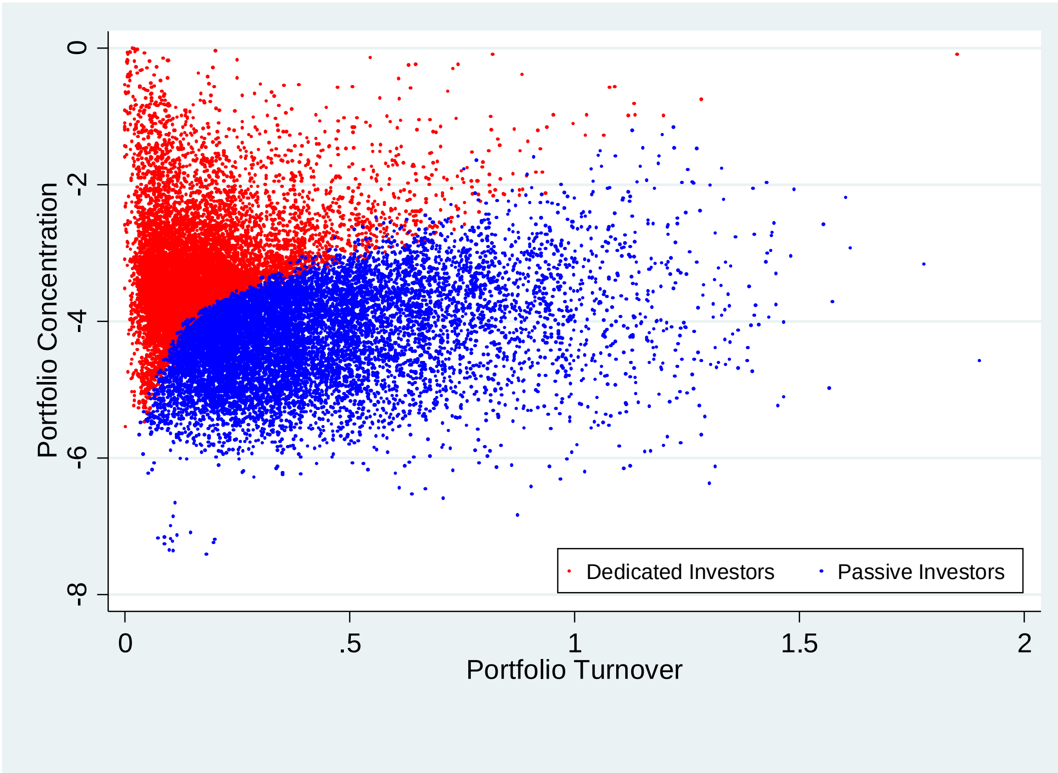

Figure 1: For 22,518 fund year observations, we group funds into (i) dedicated investors and (ii) passive investors based on a combination of their asset concentration (HHI of equity shares) and their investment horizon (proxied by the churn ratio).

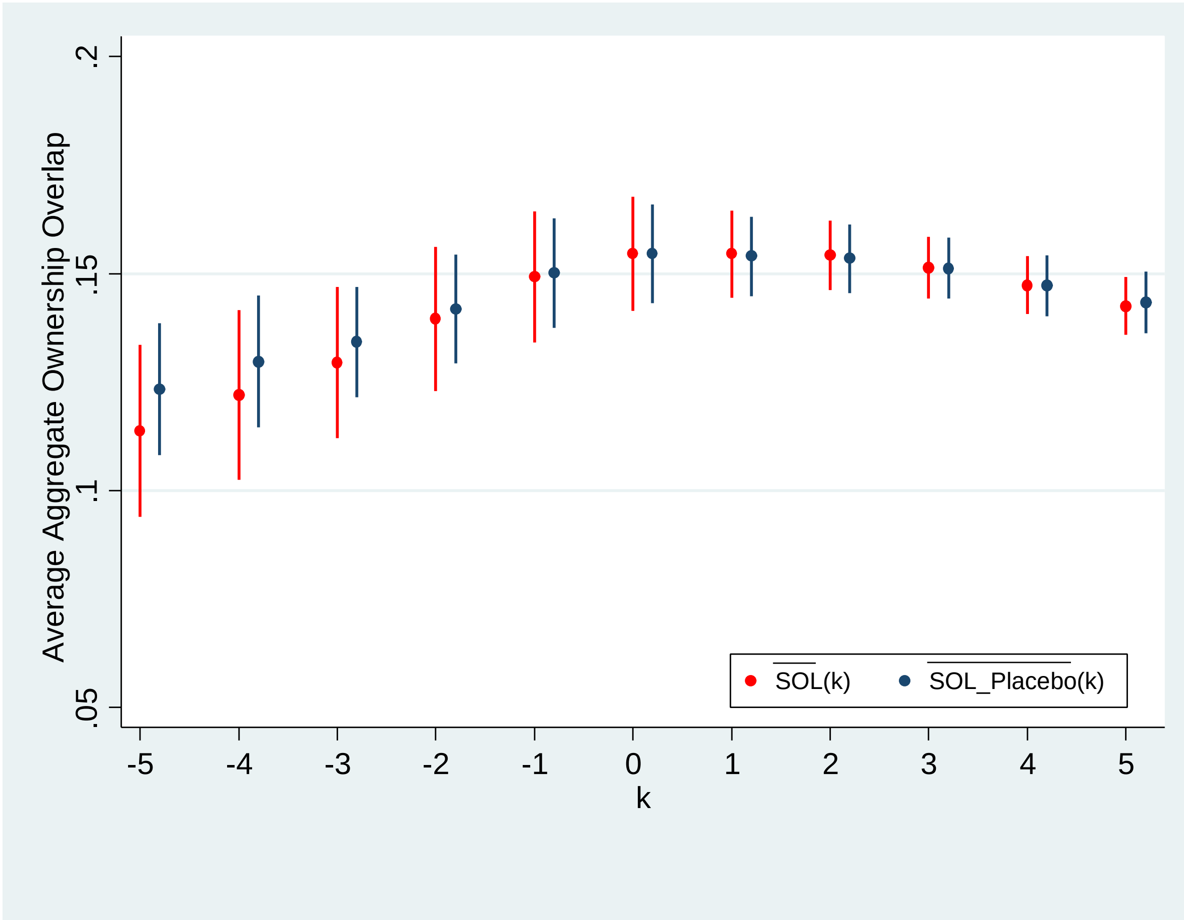

Figure 2: The evolution of the average aggregate (placebo) shareholder ownership overlap is plotted for a lag/forward of k years relative to the patent filing year ($k = 0$). The vertical lines describe a confidence interval of two standard deviations above and below the mean estimate.

Appendix

Variable	Description
$ln(1+CITES_{s,t})$	The natural logarithm of 1 plus $CITES_{s,t}$. $CITES_{s,t}$ is the number of future citations received by the cohort of patents filed by firm s in year t. We count the future citations up to the end of 2010 and exclude all self-citations. Only those patents that are ultimately granted are included in our sample. Following Hall, Jaffe, and Trajtenberg (2001), we correct for the truncation in citation count based on the estimated empirical distribution of citation-lag. [Source: NBER Patent database and Kogan et al. (2014)]
$ln(1+N_{s,t})$	The natural logarithm of 1 plus the number of patents filed by firm s in year t. Only patents that are ultimately granted are included in our sample. [Source: NBER Patent database and Kogan et al. (2014)]
$ln(1+\overline{cites}_{s,t})$	The natural logarithm of 1 plus $\overline{cites}_{s,t}$. $\overline{cites}_{s,t}$ denotes the average future citations per patent for the cohort of patents filed by firm s in year t, calculated as $CITES_{s,t}$ divided by $N_{s,t}$. [Source: NBER Patent database and Kogan et al. (2014)].
$ln(1{+}R\&D\ Exp_{s,t})$	The natural logarithm of 1 plus *R&D Expenditure* (Compustat Mnemonic: XRD), which is based on the latest fiscal year-end value prior to the end of calendar year t and is measured in million U.S. dollars. [Source: Compustat-CRSP merged database]
$ln(1+cites_{p,t})$	The natural logarithm of 1 plus $cites_{p,t}$. $cites_{p,t}$ denotes the number of future citations received by patent p, which is filed in year t. The future citations are counted up to the end of 2010, and all self-citations are excluded. The truncation bias in citation count is adjusted based on the estimated empirical distribution of citation-lag. [Source: NBER Patent database and Kogan et al. (2014)]
$PSOL(p,p_u)$	Pairwise (institutional) shareholder ownership overlap between the downstream patent p and its upstream patent p_u at the end of year t. We first identify all the overlapped (institutional) shareholders between firm s and the assignee of patent p_u. For each overlapped shareholder i, we calculate the minimum ownership overlap $min[w_{i,O(p)}, w_{i,O(p_u)}]$. $w_{i,O(p)}$ denotes the shareholding of investor i (relative to the aggregate institutional ownership) in the corporate assignee of patent p. $w_{i,O(p_u)}$ is defined analogously. Then, we calculate the sum of $min(w_{i,O(p)}, w_{i,O(p_u)})$ over all of the overlapped institutional shareholders in the two firms. When calculating $PSOL$, we ignore any upstream patent p_u whose assignee is not a publicly listed firm or whose assignee is the same as the assignee of the downstream patent p. [Source: NBER Patent database, Kogan et al. (2014), and Thomson Reuters 13F database].
$sol_{p,t}$	Shareholder overlap for patent p, filed in year t. It is calculated as the importance-weighted average $PSOL(p,p_u)$ of all cited upstream patents p_u, with $u = 1, 2, .., N_p$. We measure the importance of an upstream patent p_u by its future citations relative to the aggregate future citations of all cited upstream patents. In cases in which multiple upstream patents are assigned to the same firm, we aggregate the citation count of these patents and treat them as one single upstream patent. [Source: NBER Patent database, Kogan et al. (2014), and Thomson Reuters 13F database]
$SOL_{s,t}$	Shareholder overlap for firm s in year t. It is calculated as the importance-weighted average $sol_{p,t}$ of all patents filed by firm s in year t. We measure the importance of a patent p by its future citation count relative to the aggregate citation count of all patents filed by the firm in the year. [Source: NBER Patent database, Kogan et al. (2014), and Thomson Reuters 13F database]

Variable	Description
$SOL\_Dedicated_{s,t}$	Shareholder overlap contributed by dedicated investors. At the end of each year, we sort all institutional investors by the portfolio diversification (proxied by Herfindahl Hirschman Index (HHI)) in descending order and the portfolio turnover (proxied by churn ratio defined in Gaspar, Massa, and Matos (2005)) in ascending order, respectively, and define the combined rank as the sum of HHI rank and churn ratio rank. We label dedicated investors as those in the top half of the combined rank (high concentration and low turnover). $SOL\_Dedicated_{s,t}$ is constructed in a similar way to $SOL_{s,t}$ except that the former uses pairwise shareholder overlap from dedicated investors only. [Source: NBER Patent database, Kogan et al. (2014), and Thomson Reuters 13F database]
$SOL\_Passive_{s,t}$	Shareholder overlap contributed by passive investors. Following the definition of dedicated investors above, we label passive investors as those in the bottom half of the combined rank (low concentration and high turnover). $SOL\_Passive_{s,t}$ is constructed in a similar way to $SOL_{s,t}$ except that the former uses pairwise shareholder overlap from passive investors only. [Source: NBER Patent database, Kogan et al. (2014), and Thomson Reuters 13F database]
$SOL\_Placebo_{s,t}$	A measure for placebo shareholder overlap. For each downstream innovator, we replace every firm cited as the true upstream patent owner with a placebo firm of similar characteristics. For any firm patent cohort, the placebo firms are matched to the true upstream firms based on the same four-digit SIC industry code and then on the minimal euclidean distance of both firm size $ln(1+Assets_{s,t})$ and firm patent success $ln(1+CITES_{s,t})$, which are normalized by their respective industry average in the same year. The matched placebo firms aren't cited by any patent filed by the downstream innovator in respective year. If no matching firm is identified for an upstream firm, we then move up to three-digit SIC industry code and repeat the procedure above. [Source: NBER Patent database, Kogan et al. (2014), and Compustat-CRSP merged database]
$WHHI_{s,t}$	Weighted Herfindahl-Hirschman index of shareholder overlap concentration. First, we calculate the Herfindahl-Hirschman index of patent pair (p,p_u) for share overlap $min[w_{i,O(p)}, w_{i,O(p_u)}]$ of each investor $i \in I_{p,p_u}$, who jointly holds equity in downstream and upstream firms $O(p)$ and $O(p_u)$. $w_{i,O(p)}$ denotes the share holding of investor i (relative to the aggregate institutional ownership) in the corporate assignee of patent p. $w_{i,O(p_u)}$ is defined analogously. Second, we importance-weighted average HHI in the first step over all patent p's upstream patents p_u, with $d = 1, 2, .., N_p$, where we measure the importance of the upstream patent p_u by its future citations relative to the aggregate future citations of all patent p's peer upstream patents. In cases in which multiple upstream patents are assigned to the same firm, we aggregate the citation count of these patents and treat them as one single upstream patent. Lastly, we importance-weighted average the result obtained in the second step over all patents filed by firm s in year t, where we measure the importance of a patent p by its future citation count relative to the aggregate citation count of all patents filed by the firm in the year. [Source: NBER Patent database, Kogan et al. (2014), and Thomson Reuters 13F database]
$FIF_{s,t}$	Firm innovation focus for firm s at year t. It is the $ln(1 + CITES_{s,t})$ of firm s in year t scaled by the industry average (based on two-digit SIC codes) over the same period. [Source: NBER Patent database, Kogan et al. (2014), and Compustat-CRSP merged database]
$IIF_{i,s,t}$	Investor innovation focus for investor i of firm s at year t. It is calculated as the portfolio value-weighted average of patent share of all stocks $s\prime$ (except for stock s) invested by shareholder i. The portfolio value of shareholder i is measured at the end of year t. [Source: NBER Patent database, Kogan et al. (2014), Thomson Reuters 13F database, and Compustat-CRSP merged database]
$SIF_{s,t}$	Shareholder innovation focus for firm s at year t. It is calculated as the value-weighted average of $IIF_{i,s,t}$ across all shareholders of firm s. The weight for each shareholder i is proportional to each shareholder's investment value in the firm at the end of year t. [Source: NBER Patent database, Kogan et al. (2014), and Compustat-CRSP merged database]

Variable	Description
$IO_{s,t}$	Aggregate institutional ownership of firm s at year t. It is calculated as the total number of shares of firm s held by institutional investors relative to the total shares outstanding. [Source: Thomson Reuters 13F database and Compustat-CRSP merged database]
$IO_{s,t}^{NOL}$	Non-overlapping institutional ownership of firm s at year t. It is calculated as the total number of shares held by non-overlapping institutional investors relative to total shares outstanding. [Source: Thomson Reuters 13F database and Compustat-CRSP merged database]
$ln(1 + Assets_{s,t})$	The natural logarithm of 1 plus *Assets* (Compustat Mnemonic: AT), which is based on the latest fiscal year-end value prior to the end of calendar year t and is measured in million U.S. dollars. [Source: Compustat-CRSP merged database]
$ln(1 + R\&D\ Stock_{s,t})$	The natural logarithm of 1 plus $R\&D\ Stock_{s,t}$, where $R\&D\ Stock_{s,t} = R\&D\ Expenditure_{s,t} + (1 - \delta) \times R\&D\ Stock_{s,t-1}$. Following Hall, Jaffe, and Trajtenberg (2005), we set $delta = 0.15$ to represent the private depreciation rate of knowledge. *R&D Expenditure* (Compustat mnemonic: XRD) is based on the latest fiscal year-end value prior to the end of calendar year t and is measured in million U.S. dollars. [Source: Compustat-CRSP merged database]
$ln(1 + K/L_{s,t})$	The natural logarithm of 1 plus the ratio of *Capital* (Compustat Mnemonic: $PPEGT$) to *Labor* (Compustat Mnemonic: EMP). Both variables are based on the latest fiscal year-end values prior to the end of calendar year t. *Capital* is measured in million U.S. dollars and *Labor* in thousands. [Source: Compustat-CRSP merged database]
$ln(1 + Sales_{s,t})$	The natural logarithm of 1 plus *Sales* (Compustat Mnemonic: $SALE$), which is based on the latest fiscal year-end value prior to the end of calendar year t and is measured in million U.S. dollars. [Source: Compustat-CRSP merged database]