MANAGEMENT SCIENCE
DOI 10.1287/mnsc.1070.0814ec
pp. ec1–ec5

© 2008 INFORMS

e-companion
ONLY AVAILABLE IN ELECTRONIC FORM

Electronic Companion—"The Impact of Uncertain Intellectual Property Rights on the Market for Ideas: Evidence from Patent Grant Delays" by Joshua S. Gans, David H. Hsu, and Scott Stern, *Management Science*, DOI 10.1287/mnsc.1070.0814.

Electronic-Companion to "The Impact of Uncertain Intellectual Property Rights on the Market for Ideas: Evidence from Patent Grant Delays"

by

Joshua S. Gans, David H. Hsu *and* Scott Stern

Online Supplement

This companion to the main paper provides addition detail associated with the data, empirical framework, and empirical results of the paper.

ONLINE APPENDIX A: Data Issues

Sample Selection Criteria. We began by selecting all recorded deals in four sectors that are closely associated with cooperative commercialization between start-up innovators and more established industry players: biotechnology, electronics, software, and scientific instruments. Based on a reading of the deal description from the SDC database, we identified the first significant patent associated with the technology from searching the US Patent and Trademark Office (USPTO) website. This was done by searching patent titles and abstracts for key words taken from the SDC technology licensing activity description. This process yielded 219 patent-license pairs. By construction, our dataset excludes licenses for technologies in which no patent was ever issued, as well as technologies which are patented but never licensed. Beginning in November 2000, patent applications are disclosed 18 months after filing, as opposed to the time of patent grant (see Johnson and Popp (2003) for an analysis of the impact of the American Inventors Protection Act of 1999 (AIPA)). To impose uniformity regarding disclosure, and limit right-censoring, our sample covers the period prior to the AIPA.

While the overall analysis of deal structure across different types of players is extremely informative (e.g., Lerner and Merges, 1998), we focus our data sample in order to construct a clear test of our theoretical framework. Our sample is composed of licensing deals between start-up innovators and more established firms that are focused on specific technologies (rather than more general agreements involving long-term alliances or that are primarily focused on cross-licensing arrangements). From our initial database, we eliminate deals with the following characteristics: an established firm licensing to another established firm, an established firm licensing to a start-up, a non-profit entity as a licensor or licensee, renewal of a prior technology transfer agreement, and transactions involving strictly technology cross-licenses between or among parties. The deal was excluded if there was ambiguity over the match between the licensed technology and the patent associated with that technology, or if the licensing date was earlier than the patent application date (the latter cause for exclusion may be related to the former). This process resulted in a final sample of 198 technologies for which a patent was issued and a license was granted.

For a small set of observations (post-1999 patent grants), the HJT patent characteristics data are not available through the NBER file. We constructed the HJT measures for these observations, and checked whether our results are sensitive to their inclusion or exclusion. All qualitative results remain the same.

It is also useful to note that the industry coverage is distinct from the geography dummies in the dataset. Each of the industries is represented in each of the geographic regions (Silicon Valley, Route 128, Canada, and other), and the only significant pair-wise correlation between the industry and geography measures is a positive correlation between *Route 128* and *software* ($\rho = 0.20$).

Firm and patent characteristics. Our dataset also includes firm and patent characteristics, allowing us to evaluate the impact of observable measures of the business environment on licensing behavior. First, we define dummy variables indicating locations that may provide access to different types of technology licensing networks: *Silicon Valley*, *Route 128*, and *Canada*. As key high-technology regions, firms located in Silicon Valley and Route 128 may experience a higher overall rate of technology licensing, as network-based mechanisms may facilitate exchange even in the absence of IPR, and so licensing may be less sensitive to patent allowance in these regions. Our sample also includes a relatively high number of Canadian licensing deals, and so we construct a Canada dummy (mean = 0.18). We also include proxies for firm resources, experiences and capabilities. *Firm age* (mean = 6.03) is measured as of the patent application date, and a venture capital funding dummy, *VC funded* (mean = 0.48), only equals one for firms receiving venture funding prior to the patent application date. Access to a VC network, as well as increased maturity and reputation, might enhance the ability of a firm to engage in cooperative commercialization even in the absence of formal IPR (Hsu, 2006). Yet, firms with fewer organizational resources may be unable to delay licensing until patent allowance, and so may forego bargaining position to achieve an earlier licensing agreement. Younger firms may be less savvy in their approach to licensing, or may be willing to sacrifice bargaining power in order to quickly establish a cooperative commercialization agreement with an industry incumbent. While the overall effect of *firm age* or *VC funded* on the timing of commercialization may therefore be ambiguous, inclusion of these measures in our empirical analysis allows us to control for the possibility that differences in experience or resources may be correlated with *both* the *licensing lag* and the *patent allowance lag*.

We also incorporate several patent characteristics in the analysis. Most of these measures are simply the standard measures from the Hall et al. (HJT, 2001) NBER data file. *Patent claims* is simply the number of claims allowed by the examiner (mean = 20.84), while *patent classes* is the number of distinct primary three digit patent classes to which the patent is assigned (this measure ranges from 0-9; mean = 1.90). *Patent citations made* is equal to the number of “backward” citations to prior patents (mean = 11.17). *Patent backward citation lag* is the number of years between the *patent grant date* and the average grant year of those cited patents (mean = 7.56), and *patent originality* (mean = 0.43) measures the diversity of cited references (similar to a traditional Herfindahl index in which the measure ranges from zero to one, and is increasing in the uniformity of cited patent classes). We also include the number of non-patent references to the scientific literature (*science references*, mean = 7.56) and the number of non-patent, non-scientific references (*non-science references*, mean = 2.40). These patent characteristics may be informative about the incentives for pre- versus post-allowance licensing, such as the importance of productive efficiency, the level of tacit knowledge, or patent scope, and so may influence the baseline hazard rate of licensing, or mediate the salience of patent allowance itself. Of course, the interpretation of each measure is subtle (HJT, 2001; Lanjouw and Schankerman, 2004). *Patent citations made* may indicate a higher level of technological complexity (and therefore a higher level of tacit knowledge disclosure for effective commercialization), or alternately, a high level of this variable may be associated with significant uncertainty over the ultimate (enforceable) scope of a patent, since patent rights are more uncertain in the presence of a patent thicket (Shapiro, 2001). Similarly, while a higher level of *patent claims*, *patent classes*, or *patent originality* indicates a higher level of technological complexity and the likely importance of tacit knowledge, these measures may also be associated with increased patent scope (Lanjouw and Schankerman, 2001). While some authors argue that *science references* (and perhaps *non-science references*) indicate a higher degree of transparency for an invention (Fleming and Sorenson, 2004), Lowe (2004) suggests that patents including *science references* are more likely to require a high level of tacit knowledge exchange for effective transfer.

ONLINE APPENDIX B
Timing Lag Distributions

Figure 1A. Distribution of Patent Allowance Lag

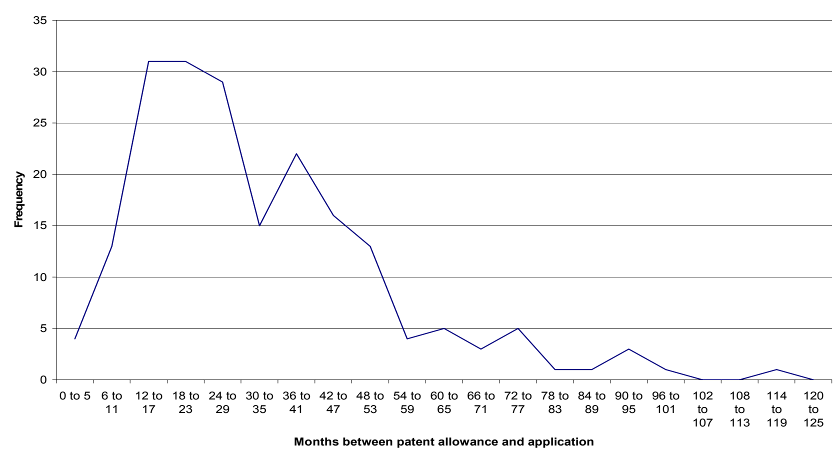
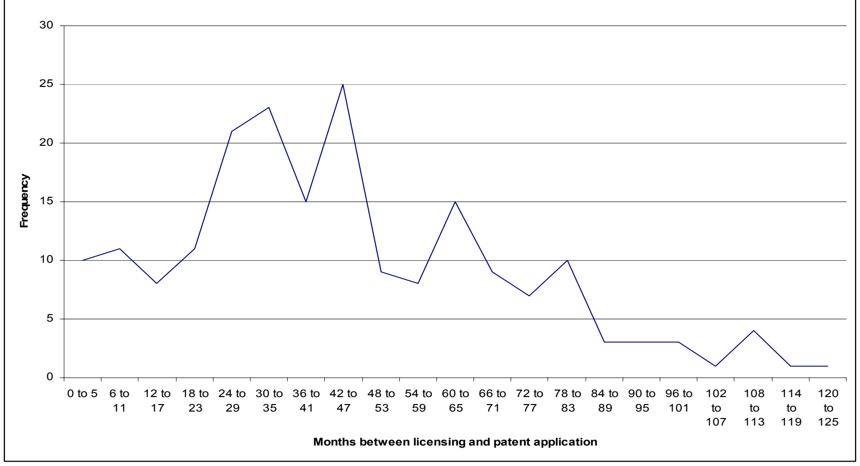


Figure 1B. Distribution of Licensing Lag



While only a very small number of technologies receive a patent allowance within a year of the application date, the majority of the technologies in our sample receive a patent allowance in the second, third, and fourth year after application. As well, the patent allowance lag has a large right tail, with a small number of technologies with patent allowance lags in excess of nine years. It is possible that extreme lags may be associated with technologies in which productive efficiency considerations may not be crucial; accordingly, we have experimented extensively with imposing a maximum patent allowance lag (e.g., 60 months). None of our key qualitative findings are affected.

In contrast to the *patent allowance lag* distribution, *licensing lag* is more evenly distributed. Figure 2 in the text of the main paper combines these histograms in reporting the distribution of *licensing lag* less *patent allowance lag*. Finally, it is useful to note that if we plot the histogram of *licensing date* less *patent grant date* (rather than *patent allowance date*), there is a pronounced increase in the rate of licensing in the four to six months prior to the *patent grant date*, which peaks in the first few months after the *patent grant date*. This is consistent with the behavior of managerial response to the event associated with uncertainty reduction (the patent allowance date) rather than the date at which formal rights commence and the patent grant is published.

ONLINE APPENDIX C: The Empirical Framework

In our discussion of the empirical framework, we discuss but do not present the specifications for the tests of our supplementary hypotheses. First, to evaluate whether licensing is "clustered" immediately after the patent allowance date, we define a set of "window" variables (*pre patent allowance* (*k,l*) and *post patent allowance* (*k,l*)), equal to 1 from *k* to *l* months prior to (or after) the *patent allowance date*, and 0 otherwise:

$$h_{LICENSE}(t, \text{POST PATENT}_i^t, l, Z_i) = h^l(t) \cdot \exp\left\{ \begin{array}{l} \beta_0 + \beta_Z Z_i + \beta_{\text{PATENTLAG}} \text{PATENT LAG}_\text{i} + \sum_{k,l} \psi_{PRE\_k,l} \text{PRE PATENT}(k,l)_i^t \\ + \sum_{k,l} \psi_{POST\_k,l} \text{POST PATENT}(k,l)_i^t + \nu_i \end{array} \right\} \quad \text{(C-1)}$$

Second, we introduce several interaction terms between *post patent allowance* and measures of the strategic and technological environment. To do so, we de-mean each element of our control vector Z_i (i.e., calculate $\bar{Z}$) to formulate the following hazard model:

$$h_{LICENSE}(t, \text{POST PATENT}_i^t, l, Z_i) = h^l(t) \cdot \exp\left\{ \begin{array}{l} \beta_0 + \beta_Z Z_i + \beta_{\text{PATENTLAG}} \text{PATENT LAG}_\text{i} + \beta_{\text{POST PATENT}} \text{POST PATENT}_i^t \\ + \beta_{PATGRANT,Z} \text{POST PATENT}_i^t \cdot (Z_i - \bar{Z}) + \nu_i \end{array} \right\} \quad \text{(C-2)}$$

This allows us to estimate the overall impact of *patent grant date* on *licensing* and how this changes with changes in the underlying economic, strategic, and technical environment.

Finally, it is possible to incorporate multiple time-varying regressors and to distinguish whether the key "shock" to the *licensing* hazard rate results from the *patent allowance date* or from the subsequent formal *patent grant date*, as follows:

$$h_{LICENSE}(t, \text{POST PATENT}_i^t, l, Z_i) = h^l(t) \cdot \exp\left\{ \begin{array}{l} \beta_0 + \beta_Z Z_i + \beta_{\text{PATENTLAG}} \text{PATENT LAG}_\text{i} + \beta_{\substack{\text{POST PATENT} \\ \text{ALLOWANCE}}} \text{POST PATENT ALLOWANCE}_i^t \\ + \beta_{\substack{\text{POST PATENT} \\ \text{GRANT}}} \text{POST PATENT GRANT}_i^t + \nu_i \end{array} \right\} \quad \text{(C-3)}$$

ONLINE APPENDIX D

Robustness to Functional Forms and Estimation Methods

Dependent Variable = LICENSE

(Robust standard errors are clustered by firm)

N = 8045

Independent Var.	**(D-1)**		**(D-2)**		**(D-3)**		**(D-4)**
	Cox proportional hazard models				*Shared gamma frailty Cox regression*		*Weibull-distributed failure time*
	Haz. Ratio	Coef.	Haz. Ratio	Coef.	Haz. Ratio	Coef.	Coef.
Post patent allowance	3.026*** (0.667)	1.107*** (0.220)	1.751** (0.474)	0.560** (0.271)	3.298*** (0.666)	1.216*** (0.202)	0.859*** (0.244)
Inverse of patent allowance lag	15.059 (31.972)	2.712 (2.123)	1.501 (4.276)	0.406 (2.849)			
Square of patent allowance lag			1.000 (0.000)	-0.000 (0.000)			
Patent allowance lag			0.994 (0.021)	-0.006 (0.022)			-0.012*** (0.004)
Patent App. Yr. FE	Yes		Yes		Yes		Yes
Log likelihood	-544.639		-537.594		-799.962		-157.177

** and *** indicates statistical significance at the 5% and 1% levels, respectively.

This table includes a number of additional empirical specifications exploring the robustness of the baseline results in Table 2. In the spirit of a control function approach, (D-1) and (D-2) include alternative functional forms for the treatment of the *patent allowance lag* (including the inverse (D-1) and the inclusion of the inverse, level and square of *patent allowance lag* (D-2)). In (D-3), we allow for "shared frailty" among technologies with similar *patent allowance lags* (we allow for 13 separate groupings based on six-month allowance lag windows and assume a gamma distribution), and in (D-4), we experiment with a specific functional form (Weibull) for the baseline hazard rate. In each of these alternative specifications, the estimated coefficient on *post patent allowance* remains large and statistically significant; indeed, the estimated impact of *post patent allowance* is actually higher for each these alternative assumptions and control structures than the coefficients reported in Table 2.