

Patents and Research Investments: Assessing the Empirical Evidence

By ERIC BUDISH, BENJAMIN N. ROIN, AND HEIDI WILLIAMS*

Competitive markets may under-incentivize private research investments relative to what the social planner would prefer. The patent system aims to address this potential under-investment problem by granting innovators a fixed time period during which they can charge supra-competitive prices, thus increasing incentives for private research by allowing innovators to capture a higher share of the social returns to their inventions.

A well-developed theoretical literature — dating back at least to Nordhaus (1969) — has analyzed optimal patent policy design. In this paper, we have three main goals. First, we re-present the core trade-off of the Nordhaus model in a manner more similar to how it would be presented today, with the aim of making the model more accessible to current readers. Second, we highlight an empirical question which emerges from the Nordhaus framework as a key input into optimal patent policy design: namely, what is the elasticity of R&D investment with respect to the patent term? Finally, we review the — surprisingly small — body of empirical evidence that has been developed on this question over the nearly half century

since the publication of Nordhaus’s book.

I. A model of optimal patent length: Nordhaus (1969)

The Nordhaus (1969) model of optimal patent length identifies the following core trade-off. On one hand, increasing patent duration benefits society by eliciting R&D activity that would otherwise not have been conducted, which yields socially valuable inventions. On the other hand, increasing patent duration harms society by giving additional monopoly protection to the inventions that society would have enjoyed even absent the increase in protection, which leads to socially harmful supra-competitive pricing. Optimal policy equates these benefits and costs at the margin.

We present a simplified and slightly modified version of the Nordhaus (1969) model, mostly following the notation of Budish, Roin and Williams (2015).¹ A representative firm conducts R&D. In the original Nordhaus model R&D is a scalar decision variable, and R&D benefits the firm and society by lowering the firm’s production costs for its single output good. In this presentation, to facilitate the discussion of empirical elasticities, the firm’s R&D decision is instead modeled as a decision over which potential R&D projects to pursue, and R&D benefits the firm and society by bringing to market inventions that otherwise would not have existed.

Potential inventions are indexed by $i \in I$, and in total have unit mass. Associated with each potential invention are: the cost,

* Budish: Booth School of Business, University of Chicago, 5807 South Woodlawn Avenue, Chicago IL 60637 (e-mail: eric.budish@chicagobooth.edu); Roin: Sloan School of Management, MIT, 50 Memorial Drive, E62-465, Cambridge MA 02142 (e-mail: broin@mit.edu); Williams: Department of Economics, MIT, 50 Memorial Drive, E52-452, Cambridge MA 02142 (e-mail: heidiw@mit.edu). We are grateful to Fiona Scott Morton and Carl Shapiro for comments, and to Sam Grondahl and Myles Wagner for excellent research assistance. Research reported in this publication was supported by the National Institute on Aging and the NIH Common Fund, Office of the NIH Director, through grant U01-AG046708 to the NBER; the content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH or NBER. Financial support from NSF grant 1151497 and the Initiative on Global Markets at Chicago Booth is also gratefully acknowledged.

¹What we refer to as the Nordhaus model is Nordhaus (1969), Chapter 5, pages 76-86. Much of the notation used in this analysis is defined and explained in Nordhaus (1969), Chapter 2. Importantly, as in the original Nordhaus (1969) book, this model and our empirical discussion abstracts away from important topics such as how patents may affect cumulative innovation.

c_i , of pursuing the invention; the probability, p_i , that the R&D will successfully yield the invention; the annual profitability, π_i , of the invention to a monopolist; the annual social value of the invention, v_i^m and v_i^c , under monopoly and competitive pricing, respectively; and the number of years of socially useful life of the invention, T_i , that is, the amount of time until the invention becomes obsolete. For simplicity, we assume that after a patent expires there is free entry and firm profits drop to zero. We also assume that the cost of pursuing the invention is a one-time cost incurred at time 0, that the R&D takes no time to conduct, and that the annual amounts π_i , v_i^m , and v_i^c grow at the discount rate, which is the same for the firm and society. Together, these assumptions let us ignore discounting which simplifies the math considerably.² Note too that in the original Nordhaus model the parameters π_i , v_i^m , and v_i^c are implicit in a demand system for the firm's product; explicitly modeling demand is unnecessary for our purposes here, but is essential in models of optimal patent breadth or models incorporating business stealing effects (as in, e.g., Gilbert and Shapiro (1990); Klemperer (1990)).

The social planner chooses t_{patent} , the number of years the firm enjoys a monopoly for an invention whose R&D is successful. Hence, the firm will choose to pursue invention i if and only if $p_i \cdot \min(t_{patent}, T_i) \cdot \pi_i \geq c_i$; that is, if the number of years of expected monopoly (the success probability times the minimum of the patent life and the total life) times per-year profitability exceed the R&D costs. Social welfare from invention i , should the firm pursue the invention, is $p_i \cdot [\min(t_{patent}, T_i) \cdot v_i^m + (T_i - \min(t_{patent}, T_i)) \cdot v_i^c] - c_i$.

What is the optimal patent term t_{patent} ? Let $\mathbf{1}_{\{\cdot\}}$ denote the indicator function which returns 1 if the statement in brackets is true and 0 if not. Let $EML_i = p_i \cdot \min(t_{patent}, T_i)$ and $ETL_i = p_i \cdot T_i$, denote the expected monopoly life and expected total life of the

invention, respectively. The optimal patent term solves the following program:

$$(1) \quad \max_{t_{patent}} \int_I \mathbf{1}_{\{EML_i \cdot \pi_i \geq c_i\}} \times \left[\underbrace{ETL_i \cdot v_i^c}_{\text{value of new inventions}} - \underbrace{EML_i \cdot (v_i^c - v_i^m)}_{\text{deadweight loss}} - c_i \right] di.$$

The solution to Equation (1) will depend on the distribution of invention parameters. To develop the intuition for the core trade-off, consider a marginal increase in the patent term t_{patent} . This has benefits and costs. The benefits are that more inventions – those that satisfy $EML_i \cdot \pi_i = c_i$ with equality – will be elicited on the margin. Let ξ denote the quantity of inventions elicited at the margin – this is the key elasticity parameter that we discuss in greater detail in Section II.³ Then the benefits from increasing the patent term at the margin can be written as:

$$(2) \quad \text{Benefits} = \underbrace{\xi}_{\substack{\text{elasticity of R\&D} \\ \text{wrt patent term}}} \times \underbrace{\mathbb{E}_{EML_i \cdot \pi_i = c_i} [ETL_i \cdot v_i^c - EML_i \cdot (v_i^c - v_i^m) - c_i]}_{\text{social value of marginal inventions}}.$$

The cost of increasing the patent term at the margin is that inventions that would have been elicited anyways – those that satisfy $EML_i \cdot \pi_i > c_i$ strictly – are given additional time on patent, which causes additional deadweight loss. These costs can be written as

$$(3) \quad \text{Costs} = \int_I \underbrace{\mathbf{1}_{\{EML_i \cdot \pi_i \geq c_i\}} \mathbf{1}_{\{T_i > t_{patent}\}}}_{\text{intensive margin}} \times \underbrace{(v_i^c - v_i^m)}_{\text{deadweight loss}} di.$$

³Note that we refer to the quantity of inventions elicited at the extensive margin from a change in the patent term as the elasticity, but in practice the empirical evidence we will review attempts to estimate the percentage increase in research investment caused by a one-year increase in the patent term.

²See Budish, Roin and Williams (2015) for a model in which both excess private discounting and R&D commercialization lags play central roles in the analysis.

Estimating the optimal patent term in practice requires estimating the components of Equations (2) and (3). Standard methodologies from fields such as public finance and industrial organization can guide the estimation of most of these components, such as the deadweight loss term and the social value of inventions. More conceptually difficult is measuring the invention elasticity ξ , which essentially requires drawing inferences about inventions that could have been developed – in the sense of being scientifically feasible – but were never brought to market because the current patent term was insufficient to incentivize their development. As we discuss in Section II, it has thus far proved difficult to construct credible counterfactuals which allow for the estimation of this key parameter.

Equations (2) and (3) also suggest some heuristic comparative statics for how optimal patent terms should vary across technologies. The benefits of a marginal increase in the patent term will be higher when R&D activity is more sensitive to changes in the patent term (that is, when ξ is larger), and when marginal inventions are of higher social value. To take a simple example, if the social value of marginal research on disease prevention is higher than the social value of marginal research on treating diseases – perhaps because firms capture a smaller portion of the social value from that disease prevention research – then society would want longer patent terms for disease prevention. The costs of a marginal increase in the patent term will be higher when a higher share of potential R&D would be conducted even in the absence of patents (i.e., there are many inventions on the intensive margin), and when the deadweight loss from increasing patent protection is larger. As a simple example of the former, if marginal software inventions are much more likely to be developed in the absence of patents than are marginal pharmaceutical inventions, then society would want longer patent terms for pharmaceuticals than for software.⁴ As a

simple example of the latter, if marginal research on disease prevention leads to inventions with low deadweight loss – for example, because there is relatively little monopoly-pricing distortion for preventatives – then society would want comparatively long patent terms for preventatives.

II. Bridging theory and data: Taking stock of the empirical evidence

A wide variety of methodologies have been used to investigate the invention elasticity ξ linking patents and research investments, including e.g. the influential line of survey work by Mansfield (1986) and others. We here focus attention on studies which have attempted to identify observational sources of variation in patent protection, and use this variation to empirically estimate the invention elasticity ξ .

A. Patent law changes as variation

A natural starting point for estimating the elasticity of R&D investment with respect to the patent term is to look for variation over time or across areas in patent laws. To the best of our knowledge, the first such contribution was Sakakibara and Branstetter (2001), who investigate how the research investments of Japanese firms responded to a set of 1988 reforms strengthening Japanese patent protection. Using a variety of datasets including a survey of firm-level R&D spending, they uncover no evidence that stronger Japanese patent rights induced higher levels of research investments among Japanese firms.

This conclusion that country-specific patent law changes induce no measurable increase in domestic R&D investment also emerged from the work of Qian (2007), who

what share of their inventions would not have been developed had patent protection been unavailable. While that survey pre-dates many now-controversial types of patents such as software and business method patents, his survey estimates would suggest that optimal patent terms would be longer for pharmaceuticals than they would be for e.g. electrical equipment, because firms self-report in that survey that a higher share of electrical equipment products would have been developed even in the absence of patent protection.

⁴This example is in the spirit of Mansfield (1986), who reports the results of a survey which asked firms

analyzed the passage of national pharmaceutical laws in 26 countries from 1978-2002; and from the work of Lerner (2009), who analyzed the impact of major patent policy shifts in 60 countries over 150 years.

Such analyses of patent law changes face several limitations. As caveated by the authors of these studies, as implemented in practice – often as part of trade negotiations – patent law changes may affect domestic R&D through other mechanisms such as changes in foreign competition. A more substantive concern is that these studies investigate how R&D investments by domestic firms respond to domestic changes in patent laws. Conceptually, this concern raises two separate issues. First, if a “large” economy like the US were to lengthen its patent term, we would expect that to affect R&D investments of non-US firms who sell products to US consumers. Second, because technologies are developed for a global market, country-specific patent law changes in “small” economies may be a relatively small source of variation in global R&D incentives. That is, a priori, one might expect to find that the change in private research investments induced by a smaller economy extending its patent term from – say – 17 years to 20 years might be quite small. This is one potential reason why the available empirical estimates of how domestic R&D investments respond to country-specific changes in patent strength may all center around zero.

One exception to this conclusion is provided by Abrams (2009), who estimates how patent filings (as a measure of R&D) responded to patent term adjustment under the Agreement on Trade Related Aspects of Intellectual Property Rights (TRIPS). Abrams estimates that a 114 day increase in patent term generates a 21% increase in patent filings, implying that a one year increase in patent term would generate a 66% increase in patent filings. However, Abrams acknowledges that this extremely large response could be driven largely or completely by substitution in when patent applications were filed over time, as his data suggest a large amount of strategic “bunching” of application filings around the date

of the policy change.⁵

B. Alternative sources of variation

Given the potential limitations associated with using patent law changes as variation, a question naturally arises of whether other types of variation could be used to estimate the elasticity of R&D investment with respect to a change in the patent term.

In Budish, Roin and Williams (2015), we investigate the following question: are private research investments distorted away from long-term research projects, i.e. projects that take a long time to complete? We assess this question in the context of cancer clinical trials, where – because those trials are generally required to show evidence that a drug improves patient survival – clinical trials for cancers with short life expectancies can be completed much faster than clinical trials for cancers with longer life expectancies. We document a variety of evidence suggesting that allowing firms to conduct shorter clinical trials would indeed increase research investments. One potential mechanism for these results is the incentive provided by the patent system: because pharmaceutical firms almost always file for patents prior to starting clinical trials, and because the patent term runs from the filing date, effective patent protection is longer for drugs that reach the market faster (by nature of requiring shorter clinical trials). Importantly, there are other plausible mechanisms through which shortening clinical trials could also increase research investments, and in Budish, Roin and Williams (2015) we do not disentangle the impact (if any) of patents. However, if we make the very strong assumption that the only mechanism through which shortening clinical trials would affect research investments is through lengthening the effective patent term, then a back-of-

⁵In addition, because Abrams uses patent filings (as opposed to a measure of “real” R&D) as his outcome variable, we would expect his estimate to exaggerate the true semielasticity of interest: lengthening the patent term makes it more attractive to file patents on existing R&D investments - because the benefits of filing increase - so more patent filings could be observed even in the absence of any change in “real” R&D investment.

the-envelope calculation based on our estimates suggests an elasticity of research investment with respect to a one year increase in the patent term of 7-24%.⁶

A strength of this approach is that – unlike the patent law change approach – we focus on a quantitatively large source of variation in the effective patent protection provided to inventors who develop different types of technologies, and we measure the research investments of essentially all firms participating in the relevant markets. However, the key limitation with this approach is that – as highlighted above – translating our estimates into estimates of how patents affect research investments requires a very strong assumption (namely, that shortening clinical trials would affect research investments only through lengthening the effective patent term). Hence, we describe this example simply as an illustration of how alternative (non-patent law) sources of variation could be used to investigate how R&D responds to changes in the patent term.

III. Conclusions

A key parameter needed to inform optimal patent policy design is the elasticity of research investments with respect to the patent term. Estimating this elasticity is conceptually difficult because it requires constructing a counterfactual in which we can infer that some scientifically feasible inventions would have been brought to market under an alternative patent policy design. Despite a near half-century of research effort, we have essentially no credible empirical evidence on this elasticity. Our goal in this paper has been to make the theoretical and empirical literature on this question more accessible in hopes of encouraging the development of novel research approaches to this topic. Given that both theory and casual evidence suggest that this elasticity is likely to vary across different types of technologies, understanding heterogeneity in the relationship between patents and research investments across industries is particularly important looking

ahead.

REFERENCES

- Abrams, David S.** 2009. “Did TRIPS spur innovation? An empirical analysis of patent duration and incentives to innovate.” *University of Pennsylvania Law Review*, 157: 1613.
- Budish, Eric, Benjamin N. Roin, and Heidi Williams.** 2015. “Do firms underinvest in long-term research? Evidence from cancer clinical trials.” *American Economic Review*, 105(7): 2044–2085.
- Gilbert, Richard, and Carl Shapiro.** 1990. “Optimal patent length and breadth.” *RAND Journal of Economics*, 21(1): 106–112.
- Klemperer, Paul.** 1990. “How broad should the scope of patent protection be?” *RAND Journal of Economics*, 21(1): 113–130.
- Lerner, Josh.** 2009. “The empirical impact of intellectual property rights on innovation: puzzles and clues.” *American Economic Review*, 99(2): 343–348.
- Mansfield, Edwin.** 1986. “Patents and innovation: An empirical study.” *Management Science*, 32(2): 173–181.
- Nordhaus, William.** 1969. *Invention, Growth, and Welfare: A Theoretical Treatment of Technological Change*. MIT Press.
- Qian, Yi.** 2007. “Do national patent laws stimulate domestic innovation in a global patenting environment? A cross-country analysis of pharmaceutical patent protection, 1978-2002.” *The Review of Economics and Statistics*, 89(3): 436–453.
- Sakakibara, Mariko, and Lee Branstetter.** 2001. “Do strong patents induce more innovation? Evidence from the 1988 Japanese patent law reforms.” *RAND Journal of Economics*, 32(1): 77–100.

⁶See the online appendix for details.