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.

* 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, 77 Massachusetts Avenue, E17-222, Cambridge MA 02139 (e-mail: heidiw@mit.edu). We are grateful 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 National Bureau of Economic Research; 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.

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, c_i, of pursuing the invention; the probability, p_i, that the R&D will successfully yield

[1] What 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 abstracts away from important topics such as how patents may affect cumulative innovation.

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., 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 $\mathbf{1}_{\{.\}}$denote the indicator function which returns 1 if the statement in brackets is true and 0 if not. Let $EPL_i = p_i \cdot \min(t_{patent}, T_i)$ and $ETL_i = p_i \cdot T_i$, denote the expected patent life and expected total life of the invention, respectively. The optimal patent term solves the following program:

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

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 $EPL_i \cdot \pi_i = c_i$ 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. Then the benefits from increasing the patent term at the margin can be written as:

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

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

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

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,

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

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 additional research on disease prevention is higher than the social value of additional research on treating diseases, 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 large. Again, as a simple example, if software inventions are much more likely to be developed in the absence of patents than are pharmaceutical inventions, then society would want longer patent terms for pharmaceuticals than for software.[3]

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

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 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 a 150 year period.[4]

[4]In a more recent contribution, Aghion, Howitt and Prantl (2013) document evidence that product market reforms which increased competition induced more innovation in countries with strong patent rights relative to countries with weak patent rights. This result is consistent with stronger patent terms encouraging research investments, but is difficult to translate into an implied elasticity of R&D investments with respect to changes in the patent term. Their estimates suggest, for example, that a one unit increase in their product market reform intensity measure increased real R&D spending by an average of 950 million US dollars more in industries in countries with strong patent protection relative

Such analyses of patent law changes face several limitations. As caveated by the authors of these studies, many recent patent law changes were implemented together with changes in trade policy, which may independently 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 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. In 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. For both reasons, it is hard to know how to interpret this empirical estimate.

to the reform's effect on R&D in industries in countries with weak patent protection. Also, as the authors note, one difficulty with this approach is that countries with strong and weak patent regimes may also differ on other dimensions as well, which is why other papers in this area have focused on studying *changes* in patent laws rather than cross-sectional differences in legal regimes.

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 re-

search 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-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%.[5]

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 from 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.

[5]See the online appendix for details.

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.

Aghion, Philippe, Peter Howitt, and Susanne Prantl. 2013. "Patent rights, product market reforms, and innovation." National Bureau of Economic Research (NBER) Working Paper No. 18854.

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.

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.

Online appendix

A1. Elasticity calculation from Budish, Roin and Williams (2015)

This section re-prints (with minimal edits) Online Appendix Section A.10 of Budish, Roin and Williams (2015), as a point of comparison for the other elasticity estimates we discuss in this paper.

As described in Section II, the empirical estimates in Budish, Roin and Williams (2015) provide an estimate of how R&D investment changes with the 5-year survival rate (our proxy for clinical trial length), $\frac{\partial(\text{R\&D investment})}{\partial(\text{5-year survival rate})}$. To translate this estimate into an estimate of how R&D investment would respond to an increase in the patent term, we would like to scale $\frac{\partial(\text{R\&D investment})}{\partial(\text{5-year survival rate})}$ by an estimate of how the patent term varies with the 5-year survival rate. In practice, we do this scaling — under a strong assumption, as detailed below — using an estimate of how a drug's commercialization lag varies with the 5-year survival rate, since one less year of commercialization lag is equivalent (under this assumption) to one additional year of patent life. By combining these estimates, we can then estimate the elasticity of interest:

$$\frac{\frac{\partial(\text{R\&D investment})}{\partial(\text{5-year survival rate})}}{\frac{\partial(\text{commercialization lag})}{\partial(\text{5-year survival rate})}} = \frac{\partial\left(\text{R\&D investment}\right)}{\partial\left(\text{commercialization lag}\right)} \approx \frac{\partial\left(\text{R\&D investment}\right)}{\partial\left(\text{patent term}\right)}$$

The conceptual problem with estimating $\frac{\partial(\text{commercialization lag})}{\partial(\text{5-year survival rate})}$ is that - by construction - we only observe clinical trial length *conditional* on a drug compound being placed in clinical trials. Because - consistent with the theoretical model presented in Budish, Roin and Williams (2015) - we document that fewer drug compounds are placed in clinical trials for patients with longer survival times, we expect selection into clinical trials to bias the relationship between patient survival and clinical trial length in the set of observed clinical trials.[6] Given this selection bias in which trials are observed in our data, we cannot obtain an unbiased empirical estimate of $\frac{\partial(\text{commercialization lag})}{\partial(\text{5-year survival rate})}$. To overcome this selection problem, we instead calibrate the relationship between commercialization lag and the 5-year survival rate using the power calculation outlined in Online Appendix Section A.9 of Budish, Roin and Williams (2015).

We can approximate our estimate of $\frac{\partial(\text{commercialization lag})}{\partial(\text{5-year survival rate})}$ with an estimate of $\frac{\partial(\text{clinical trial length})}{\partial(\text{5-year survival rate})}$, given that we expect commercialization lag to scale one-for-one with clinical trial length. In the language of the power calculation outlined in Online Appendix Section A.9 of Budish, Roin and Williams (2015), we can re-write this elasticity as:

$$\frac{\partial\left(\text{clinical trial length}\right)}{\partial\left(\text{5-year survival rate}\right)} = \frac{\partial k}{\partial \mu} = \frac{k\mu^{k-1} + Rk(1-R(1-\mu))^{k-1}}{-[\mu^k \ln \mu + (1-R(1-\mu))^k \ln(1-(R(1-\mu))]}$$

where μ is the per-period survival rate of untreated individuals, k is the number of periods of patient follow-up, and R is a constant per-period multiplicative treatment effect such that in a given period $1-\mu$ individuals die in the control group and $R(1-\mu)$ individuals

[6] As discussed in detail in Budish, Roin and Williams (2015), perhaps the most natural selection story is that firms are only willing to place a drug compound in clinical trials for patients with long expected survival times if they receive permission to use a surrogate endpoint in place of survival as an endpoint; in this case, the relationship between patient survival and clinical trial length would be biased towards zero. If we estimate this relationship in our data, we do estimate a statistically significant relationship; however, the magnitude is implausibly small, consistent with our prior that this relationship would be biased towards zero (a ten percentage point increase in the five-year survival rate is associated with a 1.5 percent increase in average clinical trial length - an increase on the order of one month).

die in the treatment group, where R is constrained such that $R(1-\mu)$ is bounded by 0 and 1.

Intuitively, μ and k come in pairs - not all μ and k will generate sufficient statistical power conditional on a given technology (R). Here, we take the two (μ, k) pairs from the examples in the introduction of Budish, Roin and Williams (2015) given that by construction these are feasible pairs (given that the trials were completed), that we know these trials looked at survival outcomes (rather than some alternative surrogate endpoints), and that these examples span different ends of the spectrum of available technologies. We assume a technology of $R = 0.8$, which translates to a 20 percent improvement in the five-year survival rate; this choice of R is arbitrary but we explore robustness to alternative values of R below. Given the assumed value of R, the two examples in the introduction of Budish, Roin and Williams (2015) can be written as:

1) Metastatic prostate cancer: 5-year survival rate of 20 percent ($\mu = 0.2$)
   - Follow-up time of 12.8 months ((12.8/12)/5 implies $k = 0.213$ units in 5-year increments)
   - Total trial length of 3 years (3/5 implies $k = 0.6$ in 5-year increments)
2) Localized prostate cancer: 5-year survival rate of 80 percent ($\mu = 0.8$)
   - Follow-up time of 9.1 years (9.1/5 implies $k = 1.82$ units in 5-year increments)
   - Total trial length of 18 years (18/5 implies $k = 3.6$ units in 5-year increments)

Plugging in these values for μ, k, and R into the above formula for $\frac{\partial k}{\partial \mu}$ gives estimates of 2.234 for metastatic prostate cancer, and 0.766 for localized prostate cancer. Those estimates are in units of 5-year increments, and multiplying them by 5 to translate them into a 1-year unit gives 11.170 and 3.827. In words, a change from 0 to 1 in the 5-year survival rate translates to between a 3.827-11.170 year increase in patient follow-up time. Therefore, we use this 3.827-11.170 range as our estimate of $\frac{\partial(\text{commercialization lag})}{\partial(\text{5-year survival rate})}$.

Our estimate from Table 2 Column (1) in Budish, Roin and Williams (2015) implies that a change from 0 to 1 in the 5-year survival rate translates into an 86.9% reduction in R&D investment. Dividing this estimate of $\frac{\partial(\text{R\&D investment})}{\partial(\text{5-year survival rate})}$ by our estimates of $\frac{\partial(\text{commercialization lag})}{\partial(\text{5-year survival rate})}$ (following the formula on the previous page) implies an estimated semi-elasticity of R&D investment with respect to a one-year change in commercialization lag of between 7.779% (based on metastatic prostate cancer; 86.9/11.170 = 7.779) and 22.707% (based on localized prostate cancer; 86.9/3.827 = 22.707).

Alternatively, we can do the same calculation using total trial length (3 and 18 years) rather than follow-up times (12.8 months and 9.1 years). Reassuringly, we obtain nearly identical estimates: 7.993% (based on metastatic prostate cancer; 86.9/10.872 = 7.993) and 23.416% (based on localized prostate cancer; 86.9/3.711 = 23.416).

We can investigate sensitivity of our estimates to different assumed values of R, the quality of the technology. A 'reasonable' range of R might be between 0.15-0.95, in which case our estimated elasticities fall between 6-54%.[7]

The above back-of-the-envelope calculation requires several assumptions, but gives some sense of magnitudes. A more important and substantive assumption is needed to conclude that our estimate of $\frac{\partial(\text{R\&D investment})}{\partial(\text{commercialization lag})}$ provides a valid estimate of $\frac{\partial(\text{R\&D investment})}{\partial(\text{patent term})}$.

[7] Our metastatic prostate cancer example - where the treatment resulted in a gain of 3.9 months on average - corresponds to $R = 0.961$, which implies elasticity estimates between 6-20%. On the other extreme Gleevec, often referenced as a "miracle" drug, is estimated to have increased the five-year survival rate from 30% to 89% - implying $R = 0.157$, and elasticity estimates between 17-54%.

Specifically, these two estimates are equivalent only if a one year reduction in commercialization lag matters only through inducing a one year increase in effective patent life, and not through other channels. This would not be the case if, for example, cost differences between short and long clinical trials are large enough to be a quantitatively important driver of R&D investment decisions.