

**From the holdings of the Department of Special Collections , Princeton University Library,
Princeton, NJ**

These documents are either in the public domain, or can only be used for educational and research purposes (Fair use) as per U.S. Copyright law (text below). By accessing this file, all users agree that their use falls in the public domain or within fair use as defined by the copyright law. We do not charge any permission or usage fees for the publication of images of material in our collections, including those provided by Princeton University Library, via our website, catalog, or directly from staff. For more information regarding copyright, credit, and citation please review the information on the Copyright, Credit and Citation Guidelines page of our website.

<https://library.princeton.edu/special-collections/policies/copyright-credit-and-citation-guidelines>. If you have any questions, including how to order higher-quality images from the collection, contact Special Collections Public Services via our Ask Us! Form. <https://library.princeton.edu/special-collections/ask-us>

Doctoral Dissertations, 2011-2026

Economics

2017

Essays in Industry Dynamics and Innovation: A Computational Approach

Katzwer, Richard Moshe

Princeton, NJ : Princeton University

<http://arks.princeton.edu/ark:/88435/dsp01x346d6680>

U.S. Copyright law: The copyright law of the United States (Title 17, §108, United States Code) governs the making of photocopies or other reproductions of copyrighted material. Under certain conditions specified in the law, libraries and archives are authorized to furnish a photocopy or other reproduction. One of these specified conditions is that the photocopy or reproduction is not to be used for any purpose other than private study, scholarship, or research. If a user makes a request for, or later uses, a photocopy or reproduction for purposes in excess of fair use that user may be liable for copyright infringement. This institution reserves the right to refuse to accept a copying order if, in its judgment, fulfillment of the order would involve a violation of copyright law. <https://www.copyright.gov/title17/title17.pdf>

ESSAYS IN INDUSTRY DYNAMICS AND INNOVATION: A COMPUTATIONAL APPROACH

RICHARD MOSHE KATZWER

A DISSERTATION
PRESENTED TO THE FACULTY
OF PRINCETON UNIVERSITY
IN CANDIDACY FOR THE DEGREE
OF DOCTOR OF PHILOSOPHY

RECOMMENDED FOR ACCEPTANCE
BY THE DEPARTMENT OF
ECONOMICS
ADVISER: PROFESSOR JAN DE LOECKER

APRIL 2017

© Copyright by Richard Moshe Katzwer, 2017.

All rights reserved.

Abstract

This is a collection of essays related to dynamic economics on the boundary between Industrial Organization and microeconomics. Chapter one examines optimal patent design, and constructs a dynamic oligopoly model in which firms innovate along a quality ladder. Firms make decisions about both R&D investment as well as litigation spending, both of which drive profits. This is the first paper which studies patent policy in a dynamic setting that includes endogenous market structure and patent litigation. The model incorporates three observations about patent policy: (1) patent protection can have pro- and anti-competitive effects in a dynamic model; (2) private and social litigation costs are *not* second-order and should be modeled along with R&D; (3) policy mix matters and policy instruments cannot be studied independently. Industries characterized by higher variable R&D costs should have lower patent duration and stronger forward exclusion. Forward exclusion should be constant and patent duration should be decreasing in the market size for new innovations.

The second chapter develops new computational methods for the computation of the Abreu-Pearce-Stachetti equilibrium payoff set in games with imperfect monitoring. These games are of great interest, but are generally intractable computationally. The algorithm uses a grid search approach to plotting out the contours of the APS set. Theoretically, the proposed algorithm improves on the standard Judd-Yeltekin-Conklin algorithm by giving precise error bands on the difference between the computed approximate APS operator and the true APS operator, and by being more easily generalizable to more than two players. Testing show speed advantages for the grid algorithm in practical applications.

The final chapter returns to the patent question, and studies optimal patent duration in a dynamic oligopoly. The innovation process distinguishes between product quality innovations and process innovations, which have different implications for patent enforcement. I discuss how optimal patent duration, T^* and industry concentration $N(T^*)$ vary by industry. Product lines with high entry/fixed costs should have lengthy patent durations and small firm concentrations. Those

with rapid innovation rates or high probabilities of breakthroughs generally warrant shorter patent duration and lower market concentration.

Acknowledgements

I will forever be grateful to my adviser, Jan De Loecker. Jan has been endlessly encouraging of my research. He has always supported me when I've attempted big questions, or taken risky approaches. His infectious optimism has helped me both with the job market as well as overcoming blockages in my own research. Jan always kept me focused on answering economic questions that matter, and pulled me back from my own tendencies towards over-complication and obsessing over modeling detail. Jan helped me see the forest through the trees. I am grateful to him for the faith he put in me when he brought me to conferences, helped me to network with other Industrial Organization economists, and his incredible efforts on my behalf towards placement.

I am also indebted to Dilip Abreu for supporting me. More than anyone else, Dilip forced me to think carefully about *mechanisms* and what features are actually driving the results in a model. His sharpness often helped me to clarify and understand my own research. Dilip showed me how the individual trees in the forest are important. I am grateful to Myrto Kalouptsidi, for her excellent assistance in shaping, framing and presenting my research. Myrto's genuine interest in my models and approach was tremendously reassuring and encouraging. I would also like to acknowledge Benjamin Brooks, who in many ways served as a mentor to me when I was a new graduate student. Ben sharpened all of my skills: as a researcher, as an economist, as a programmer and as a team-member. I am thankful for all of the time he spent teaching me as we worked together on computational projects.

For offering valuable feedback and criticism during my talks in the Industrial Organization seminar, I'd like to thank Professors Jakub Kastl and Bo Honore. I'd also like to thank the student participants, namely Dan Goetz, Channing Verbeck, Giulia Brancaccio, Chang Sun, Neel Sukatme, Juha Tolvanen, and Olivier Darmouni. I am grateful to Professors Otto Toivanen, Matthew Mitchell, Jeffrey Campbell, Volcker Nocke, and Robin Lee for excellent feedback and discussant work on sections of this dissertation. Their insights have been invaluable.

I thank Laura Hedden and Gabriella G. Herrera, for their wonderful administrative support. I also gratefully acknowledge financial support from Princeton University, and the Industrial Organi-

zation group at Princeton. I would also like to acknowledge Gregory Crawford and the organizers of the Sixteenth CEPR/JIE School and Conference on Applied Industrial Organization.

The thing that will last with me the longest are the friendships that I made with my colleagues during my six years here. Six years of studying, doing problem sets, taking generals, presenting and researching forms a lasting bond. I'd like to thank Pauline Leung, Mark Li, Pedro Olea, Michael Sockin, Ricardo Reyes-Heroles and Gabriel Tenorio for their friendship. Princeton summer would not be the same without Princeton summer softball, and I'd like to thank my co-captains on the economics department team, Tom Winberry and Christian vom Lehn. Most of all, I would especially like to thank my former housemates: Sharon Traiberman, Inessa Liskovich and Diane Alexander. My housemates challenged and encouraged me to be my best, and I will never forget our many late-night discussions about economics, politics and policy.

I'd like to thank my fiancé, Rebecca, for always pushing me to overcome obstacles and succeed. Building a life with her has been the most wonderful time of my life. Lastly, I would like to thank my family: my parents, Norman and Tamar and my sister, Yael. Yael has been a constant friend through my life. Without the decades of hard work by my parents, and the values of hard work and intellectual curiosity that they instilled in me, I would not have had the life opportunities or discipline to complete this dissertation.

Dedicated to my parents, Norman and Tamar Katzwer.

Contents

Abstract	iii
Acknowledgements	v
1 Dynamic Oligopoly and Optimal Patent Design	1
1.1 Introduction	1
1.2 What are patents?	7
1.3 Model	9
1.3.1 Production and Consumption	11
1.3.2 Government	15
1.3.3 R&D Firms	16
1.3.3.1 The Litigation Subgame	17
1.3.3.2 Research and Development	20
1.3.3.3 Patent Obsolescence and Patent Renewal	23
1.3.3.4 Entry Subgame	24
1.3.3.5 Survival (or Exit) Subgame	26
1.3.4 Social Welfare	28
1.3.5 Equilibrium	29
1.4 Computation and Comparative Statics	30
1.4.1 Example Industry Dynamics	31
1.4.2 Comparative Statics	40
1.4.2.1 Static vs. Dynamic and Optimal Patent Duration	40

1.4.2.2	Optimal Patent Duration and Forward Protection	44
1.5	Conclusion	49
2	A Grid Approximation Algorithm for Solving Imperfect Monitoring Games	51
2.1	Introduction	51
2.2	Algorithm	54
2.2.1	Theoretical background	54
2.2.2	The Judd-Yeltekin-Conklin (JYC) Algorithm	56
2.2.2.1	Problems with the JYC Algorithm	58
2.2.2.2	The Grid Approximation Algorithm	59
2.3	Results	65
2.3.1	Perfect Monitoring Games	66
2.3.2	Imperfect Monitoring Games	70
2.4	Conclusion	75
3	Optimal Patent Duration in a Dynamic Oligopoly	76
3.1	Introduction	76
3.2	The Model	85
3.2.1	The Innovation Process	87
3.2.1.1	The Quality Ladder	87
3.2.1.2	Patent Policy	88
3.2.1.3	The State	90
3.2.2	Households	91
3.2.3	Firms	91
3.2.3.1	Oligopoly and Monopoly in the Stage Markets	91
3.2.3.2	Investment and Innovation	92
3.2.3.3	Remark on Entry: Zero Economic Profit Condition	93
3.2.4	The Firm's Problem	94

3.2.5	Solution Concept: Symmetric Markov-Perfect Equilibrium	94
3.3	Numerical Results	96
3.3.1	Policy Functions for fixed N and T	97
3.3.2	Finding $N(T)$ and T^*	100
3.3.3	Comparative Statics	104
3.4	The Model Going Forward	109
3.4.1	Full Ericson-Pakes Entry/Exit Dynamics	109
3.4.2	Empirical Analysis	110
3.5	Conclusion	112
A	Appendices for Ch. 1	126
A.1	The Static Model in 1.4.2.1	126
A.2	Baseline Functional Specification in Section 1.4.1	128
B	Appendices for Ch. 3	130
B.1	Utility Specifications in 3	130
B.1.1	Utility	130
B.1.1.1	Utility Specification #1	130
B.1.1.2	Utility Specification #2 (Money in the Utility Function)	132
B.1.2	Pricing and Indirect Utility	132
B.1.2.1	Utility Specification #1	132
B.2	The State Transition Probabilities in 3	134
B.2.1	Transition Probabilities Facing a Firm in a Symmetric MPE	137
B.3	Optimal Pure Strategy Investment in 3	139
B.4	Existence of a Symmetric Markov-Perfect Equilibrium in 3	141
B.5	Free Entry Matters in 3	143
B.6	Using the <code>optimalpatents</code> package from 3	143
B.7	Comparative Statics Tables in 3	147

Chapter 1

Dynamic Oligopoly and Optimal Patent Design

1.1 Introduction¹

Thomas Edison² presciently remarked that a patent was merely an “invitation to a lawsuit” (Scotchmer [1996]). In 2011, Apple and Google spent more money on patent litigation and defensive patent acquisition than they did on research and development of new goods (Duhigg and Lohr [2012]). In fact, the number of patent lawsuits filed jumped from 2,500 in 2009 to over 6,000 by 2013 (Mullin [2014]).³ In light of this phenomenon, there is broad agreement in the media and in policy circles that patent policy needs reform. In 2011, the Congress passed the America Invents Act with the stated goal of reducing costly patent litigation, and further patent reform remains a policy objective of the Obama administration (Office of the Press Secretary [2014], Executive Office

¹This chapter was previously presented at the 13th International Industrial Organization Conference Rising Stars Session (Boston, USA, April 2015) and the 16th CEPR-JIE School and Conference on Applied Industrial Organization (Zurich, Switzerland, May 2015).

²who held approximately 1,000 patents in the late 19th century.

³It is not just the number of patent lawsuits that has increased, median and maximum damages awarded in patent infringement lawsuits have also grown Byrd and Howard [2014], Howard [2015]. Between 2012 and 2013 average damages increased from \$27m to \$35m, and median damages increases from \$1.02m to \$1.26m. In 2013, six patent cases resulted in damages over \$100m. Total damages awarded in 2014 totaled \$2.1bn. Despite Apple being one of the preeminent patent litigators, Apple CEO Tim Cook stated that “there’s some of this that is maddening. It’s a waste; it’s a time suck.” Duhigg and Lohr [2012].

of the President [2014]). Patents incentivize innovation by offering rewards but can also potentially discourage cumulative innovation by providing infringement protection to past innovators.⁴ These competing effects can lead to a muddled policy debate. President Obama simultaneously praises the number of patents issued as a measure of national success, stating that

America still has the largest, most prosperous economy in the world. ... No country...grants more patents to inventors and entrepreneurs

while noting in 2014 that America issues too many patents, stating that

One of the biggest problems...is how do I deal with these folks who basically are filing phony patents and are costing some of my best innovators tons of money in court (Hatter [2014]).

This encapsulates the problem faced by policymakers: How does society calibrate patent protection in order to efficiently incentivize costly innovation?⁵

The objective of this paper is to study the optimal design of patent policy in different industries, which may vary in underlying market sizes, the social value of innovations, and the costs of developing innovations. (Optimality is defined as welfare maximizing.) For example, software and pharmaceuticals differ greatly in the fixed costs of innovation and the social value of innovations, yet both are governed by the same patent regime.

To that end, I construct a dynamic oligopoly model that realistically models strategic firm interactions in both R&D as well as litigation. I model a good which can be purchased in a variety of qualities. Consumers prefer higher qualities to lower qualities, and the quality frontier is advanced

⁴Since a key objective of patents is to encourage technology disclosure (making innovation easier), cumulative innovation seems to be the relevant type to discuss.

⁵Several economists have weighed in on the issue in the press. For example, Eric Maskin writes that in industries characterized by highly sequential progress (in which each innovation builds on the last), “it may be better for society to scrap patents altogether than try to tighten them” Maskin [2012]. marginalrevolution.com popular economist Alex Tabarrok argues the prevailing view in many circles: patent protection is excessively strong and we are on the wrong side of a patent-strength/innovation-rate “Laffer curve” Tabarrok [2012]. Michele Boldrin and David K. Levine make an argument for the wholesale elimination of patents in *Against Intellectual Monopoly*, a book intended for non-specialists Boldrin and Levine [2010b].

through the costly research activities of firms. Firms' innovations are protected by potentially overlapping patents, and the right to actually bring a given patented technology to the market (and earn profits) is resolved through the firms' litigation activity in the courts. The model is reminiscent of both the standard quality ladder models in trade and macroeconomics (Grossman and Helpman [1991]) and the dynamic oligopoly framework common in industrial organization (Ericson and Pakes [1995a], Doraszelski and Pakes [2007]). In this paper, I focus on two policy instruments, the well-defined *statutory patent duration* as well as *forward exclusion*, a measure of the leading breadth given to past innovators over future innovators.⁶ To my knowledge, this is the first paper which studies welfare-maximizing patent policy in a dynamic oligopoly framework with finite patent duration. Moreover, as far as I know this is the first paper to model the socially costly⁷ litigation component of dynamic innovation. The theoretical richness of the model comes at the cost of analytical tractability, and consequently I solve for Markov perfect equilibria numerically. Using the computed equilibria, I simulate the model to compute welfare.

The model is constructed to address three observations about patent policy. Firstly, since innovation is often cumulative, patent policy cannot be evaluated as a problem in which the government seeks to incentivize a single innovation, statically, in isolation. Innovations build on one another, and as a consequence patent protection has both pro- and anti-competitive effects. (It rewards innovators today but potentially punishes innovators tomorrow.) A fully dynamic model is necessary to accurately model these pro- and anti-competitive effects when calibrating policy.⁸

Secondly, litigation matters. The private and social costs of patent litigation are not second-order⁹, and should be formally modeled in the firm's problem and in any welfare analysis of patent policy.

⁶One cannot examine litigation (and would miss a large part of the current policy debate) without including some measure of forward exclusion in the analysis.

⁷Bessen and Meurer have a series of papers estimating the private and social costs of patent litigation Bessen and Meurer [2014], Bessen et al. [2011]. They document an explosion of patent litigation over the last few years and estimate a social cost of \$83bn a year.

⁸This point is made concretely in figure 1.19, which shows how the static on optimal patent duration flips in a when we go from a static model to the fully dynamic model.

⁹Bessen and Meurer [2014] put the direct legal costs of lawsuits from non-practicing entities alone at \$29bn per year (against an annual R&D budget of \$247bn per year). Some industry estimates I have seen put this number at a lower (but still substantial) \$12.2bn in 2014 and \$12.5bn in 2013 (RPX [2014]).

Lastly, the policy mix matters when discussing optimal patent design. It is difficult (and potentially impossible) to discuss “patent strength” abstractly without being precise on the specific patent policy instruments that are being considered as “strength”. Patent “strength” is actually a complicated function of several different policy instruments set by different regulatory agents: statutory patent duration, forward infringement protection, patent breadth, patentability and the patent maintenance fee structure.¹⁰¹¹ The paper demonstrates that these instruments do not trade off in a simple manner and must be jointly optimized by the social planner.

Instead of a “one-size-fits-all”¹² solution, this paper computes qualitative rules detailing how optimal policy should vary by industry. I find that conditional on the same market-value of quality innovations, industries with higher research costs should optimally have higher market concentration and patents that are “short” but “broad”. That is, firms in industries where innovation is more costly should be given patents with relatively smaller statutory duration but much stronger forward protection. The idea is that when innovation is very costly, it is better to have smaller market concentration to avoid duplicated innovation,¹³ wasted R&D spending and socially inefficient litigation spending.¹⁴ This is achieved by high forward exclusion, thereby reducing future competition.

Holding the costs of drawing quality innovations fixed, industries which are characterized by larger social value of innovations should have smaller patent duration but relatively constant forward protection for past innovators. That is to say, larger market profits from an innovation efficiently reward innovators and imply less need for patent protection.¹⁵ (It is not obvious that policy

¹⁰Figure 1.21 demonstrates that even if we wanted to collapse these multiple policy instruments into a single, theoretical “patent strength” index, we couldn’t. They do not trade off one-for-one. There is no way to fix optimal “patent strength” and define duration as a function of forward exclusion: both must be set optimally.

¹¹Moreover, the set of instruments which are optimized can change the recommended policy direction. Patent duration in isolation should increase for costlier industries. When set jointly forward exclusion, it should decrease for costlier industries.

¹²Scotchmer [2006] identifies two common criticisms from economists regarding the “one-size-fits-all” approach. She writes that the courts and the Patent and Trademark Office “have no mandate to make decisions based on economic reasoning. It is almost inevitable that some classes of innovations are under-rewarded and others are over-rewarded relative to the cost of innovation”. A second criticism is that “the reward structure cannot be modified according to the market structure in which the innovator operates”.

¹³Recall, we are in a quality ladder.

¹⁴In the model, litigation entails private and social court costs.

¹⁵Note though that we do not allow for infringement or imitation in production.

would go in this direction: greater market value for an innovation might imply increased patent protection to make innovation more rapid.)

Holding costs and market values of innovations fixed, industries characterized by high rates of exogenous patent destruction¹⁶ should have relatively constant patent duration but stronger forward protection. This is intuitive: when patents exogenously have higher rates of obsolescence and *effective* patent duration is shorter, it is easier to reward innovators through stronger exclusion next period rather than with lengthened statutory duration (which may not be effectively realized).

This paper addresses a question that has primarily been studied in the microeconomic theory literature using tools from industrial organization. Starting in the 1960s, economists pioneered models of optimal patent policy for a single, static innovation (Nordhaus [1967], Tandon [1982], Gallini [1992], Gilbert and Shapiro [1990], Klemperer [1990]). At first these models only considered duration, but then began allowing for other instruments such as breadth. Unfortunately, there are no clear cut takeaways from the single-innovation optimal patent literature (Rockett [2010]). Optimal patents are generally “long-and-narrow” or “short-and-broad”¹⁷, but the optimal patent is highly dependent on functional assumptions. On the other hand, my model generates interior solutions for the various instruments which is potentially more useful for policy making. In the 1990s, economists began to develop two-step models of cumulative innovation under patents (Green and Scotchmer [1995], Chang [1995], Scotchmer [1996], Denicolo [2000], Hopenhayn et al. [2006], Hopenhayn and Mitchell [2001, 2010]). These models were a great first step at capturing the effects patents have on follow-on innovation, but without full dynamics these effects won’t be properly quantified. Moreover, these papers primarily study how to efficiently split a monopolist rent between an innovator and a user of that innovation in order to incentivize the initial innovation: this is a different problem than studying the welfare effects of patent protection on cumulative innovation.

More recently there have been several papers which study patents in a dynamic setting. None of these papers model the large private and social costs of patent litigation in their study of dynamic

¹⁶The value of patents can be exogenously “destroyed” for a host of reasons: changing market taste, paradigm-shifting innovation in a related field, etc.

¹⁷Often bang-bang, zero or infinity.

innovation. A seminal contribution is O'Donoghue et al. [1998], which compares two patent regimens: Infinite duration and narrow breadth (so effective patent life is finite due to new innovations) and small duration and broad breadth (so effective patent duration equals the statutory duration). In O'Donoghue et al. [1998], firms decide whether to “purchase” innovations that arrive exogenously. In my model, it is truly the R&D efforts of firms that drive innovation. The two papers most similar to my analysis are Hunt [2004] and Horowitz and Lai [1996]. Both papers involve assumptions that make them unsuitable for modeling litigation and cumulative innovation. Hunt [2004] builds a dynamic oligopoly model with sequential innovation, but focuses on “patentability”¹⁸ rather than statutory patent duration. He assumes a *passive incumbent*, that is patent holders do not try to innovate.¹⁹ Hunt derives an optimal innovation step below which innovations cannot be patented.²⁰ Patents are automatically voided by subsequent innovation, whereas in my model patents are not automatically voided, they must be litigated. The assumption of a passive incumbent seems too unrealistic when we are trying to study patent litigation, in which litigants are frequently innovators both today as well as in the past.²¹

Horowitz and Lai [1996] derives a welfare-maximizing finite patent duration in a dynamic quality ladder model. They show that the welfare-maximizing patent duration is less than the rate-of-innovation-maximizing patent duration.²² The key difference between Horowitz and Lai [1996] and this paper is that Horowitz and Lai assume a *single* innovator who only faces competition

¹⁸O'Donoghue [1998] discusses patentability in depth. Similar in some sense to leading breadth, patentability specifies the minimum improvement over existing qualities that can earn a patent. This can be shorter or longer than leading breadth. If I am in a standard quality ladder, the leading breadth may be of size Δ_{LB} , so that if quality μ is the patented state of the art, then qualities $\tilde{\mu} \in [\mu, \mu + \Delta_{LB}]$ infringe on the μ -patent if they are produced. A separate instrument, patentability Δ_P may state that only qualities $\tilde{\mu} \geq \mu + \Delta_P$ can be patented as sufficiently original. If $\Delta_{LB} < \Delta_P$, qualities $\tilde{\mu} \in (\mu + \Delta_{LB}, \mu + \Delta_P)$ can be produced by competitors to the patent-holder, but not patented.

¹⁹This assumption is an extreme limit case of the endogenous result that I find, that patent holders will innovate less than their challengers. Hunt concedes that the assumption is entirely *ad-hoc*.

²⁰Unlike most papers in this literature, Hunt has an endogenous industry structure which is appealing. The fact that effective patent duration is solely a function of innovation success rates and not any exogenous duration schedule removes time as a state variable and makes the analysis far simpler than models in which the clock on patent duration is winding down. It is this construction which makes the value functions simple and allows him to solve the model analytically; his methods would not work in a model with statutory patent duration.

²¹Hunt solves for an innovation-rate-maximizing patent policy, which need not be welfare maximizing.

²²In my model, if I fix the number of firms I can also replicate the result that the welfare-maximizing patent duration is usually less than the innovation-rate-maximizing patent duration. In my paper, the patent duration that maximizes $\mathbf{E}[\mu_{t+1} - \mu_t \mid T]$ is best. This disappears when one allows for firm entry. The single innovator in their model behaves in qualitatively similar ways to the incumbent patent holder in my model.

in production and *not* in research and development. This paper has an endogenous number of potential innovators. Horowitz and Lai [1996] cannot speak to the issue of forward breadth because of the monopolist research firm.

This paper joins a growing literature in Industrial Organization in which theoretical models (that defy analytical solution) are solved numerically in order to answer complex policy and behavioral questions.²³ Some questions studied using a numerical approach include collusion and price wars (Fershtman and Pakes [2000]), learning and forgetting in dynamic competition (Besanko et al. [2010]), technology adoption (Doraszelski [2004]), and optimal merger policy (Mermelstein et al. [2014]).

The rest of the paper is organized as follows: Section 1.2 defines patents. Section 1.3 details the model. I outline the innovation process, how patents are defined in the model, the consumer's problem, the firm's dynamic R&D investment problem, as well as the structure of litigation. Section 1.3.5 defines equilibrium. In section 1.4, I numerically solve for equilibria. I explore the properties of a sample equilibrium for a baseline specification. Next, I provide some comparative statics on cost and demand primitives, and discuss how optimal policy varies in the primitives of the model. Section 1.5 concludes, with mention of avenues of future research.

1.2 What are patents?

A patent is a form of intellectual property on an innovation or invention. A patent provides “a right...to exclude others from making, using, or selling” the product or process under patent Rockett [2010]. Under U.S. patent law, many different types of innovations can be patented; new molecules (pharmaceuticals), an industrial technique or process, a new mechanical invention, an improvement on an old mechanical invention and even certain shapes in industrial design. A key distinction must be made between “right to exclude” and “right to practice”. A patent on A gives a firm the right to exclude others from producing A , but does not ensure that the firm can produce A

²³This is sometimes called “theory with numbers”.

itself. For example, firm 1 could patent the technology with components *ABC* while firm 2 could own the patent to *A*. Firm 1 does not have a right to practice *ABC*. Right to practice is resolved either through litigation or licensing.

Bessen and Maskin [2009] state that the standard economic rationale for intellectual property is to protect innovators from competitor imitation in order to incentivize them to incur the costs of research.²⁴ ²⁵ Through patent protection, patent holders are granted temporary, artificial monopolist rents leading to deadweight loss and reduced consumer surplus.²⁶ Patent protection promotes innovation that would otherwise not occur, or not be disclosed.

Patents are characterized by several policy instruments: Patent duration refers to the number of years that a patent holder is granted monopolist rights to exercise their patent.²⁷ Note that they don't have an automatic right to produce the innovation, only to block potentially infringing production by others. *Effective* patent duration is what matters; the number of periods the patent holder can actually profit from the patent they hold. This need not be the same as the *statutory* patent duration that is legislated.²⁸

Lagging patent breadth refers to the size of the “neighborhood” of products around the innovation that are deemed to infringe on the patent and cannot be produced/used.²⁹ Forward patent

²⁴This is the patent justification I focus on this paper. A second justification is that the patent system provides an organized disclosure system of new information, in which the new knowledge (if not the right to profit from it) becomes a public good. The ability to act on the knowledge is granted when the patent expires. (Lawyers refer to this as the “teaching” provided by the patent Scotchmer [2006].) Boldrin and Levine [2012] demolish this argument, writing that “from a theoretical point of view... [this justification] fails even the simplest model”. They observe that if patent protection is m periods, then only the innovations that can be kept secret for $n \leq m$ periods will be patented, with the remainder not being patented. So there is no net disclosure gain from a patent system.

²⁵Somewhat uniquely, the U.S. Constitution recognizes the right of innovators to profit from their innovation, writing in Article 1, Section 8 that a role of Congress is

To promote the Progress of Science and useful Arts, by securing for limited Times to Authors and Inventors the exclusive Right to their respective Writings and Discoveries.

²⁶Legally, “term of patent”. In the United States, patent duration is twenty years from the date of filing, and is legislated in Title 35 Section 154 of the U.S. Code, “Contents and term of patent; provisional rights”. See: <http://www.uspto.gov/web/offices/pac/mpep/s2701.html#sect2701>

²⁷Interestingly, the U.S. Patent Act of 1790 allowed for patent duration to be determined on a case-by-case basis, but not to exceed fourteen years. See: http://www.ipmall.info/hosted_resources/lipa/patents/Patent_Act_of_1790.pdf

²⁸For example granting an infinite statutory patent duration on vacuum tubes has an effective patent duration of zero periods as the technology is obsolete.

²⁹One can differentiate between “lagging” and “leading” breadth; lagging breadth refers to product qualities below the state of the art that are too similar to the state of the art and cannot be produced/used. Leading breadth refers

breadth refers to the protection given to patent holders from follow-on innovators. Forward breadth is often taken to be identical to *patentability*, the minimum size of an innovation that can earn a new patent from the patent authorities. However, it is perhaps more relevant to define forward breadth as some metric of how generous the *courts* are to past innovators when evaluating patent infringement claims against new innovators. Naturally, patent breadth and statutory patent duration jointly determine effective patent duration.

Other policy instruments include the patent maintenance fee schedule (the fees paid by patentees to keep a patent active), and more indirectly cost R&D subsidization in fields that are patent-heavy.

The United States is characterized by a “one-size-fits-all” patent policy.³⁰ Software development and bicycle manufacture are industries characterized by different rates of innovation, cost structures, entry barriers, etc., but the same patent policy is applied to innovations in both.³¹ Patent duration is defined through legislation (and is invariant across industries), and patent breadth is subjectively defined on a case-by-case basis by the U.S. Patent and Trademark Office and by the courts.

1.3 Model

I develop a dynamic model of research, production and consumption in a market for a single “good” which comes in a variety of qualities. Consumers prefer higher qualities to lower qualities. The time- t frontier quality of the good, μ_t , advances stochastically according to the research and development activity of firms. Quality improvements are rewarded with patents. The right-to-

to new, superior qualities that are too similar to the state of the art to be patented or used without infringing on the existing patent. A seminal paper on patent breadth is O’Donoghue [1998]. O’Donoghue calculates an optimal “leading breadth”, that is the the quality improvement on the existing state of the art necessary to earn a new patent. He finds that the optimal leading breadth is positive. O’Donoghue concedes the inherent unquantifiability of “patent breadth”. In my model, breakthroughs ($\mu_{t+1} - \mu_t = \Delta$) are effectively granted infinite patent breadth for the duration of the patent.

³⁰With the notable exception of additional FDA requirements for pharmaceuticals.

³¹Software design (including pseudo-code algorithm descriptions) are patentable;
http://en.wikipedia.org/wiki/Software_patents_under_United_States_patent_law

practice is resolved through litigation. Market structure is endogenous. The structure of a typical period of the model is outlined in diagram 1.1.

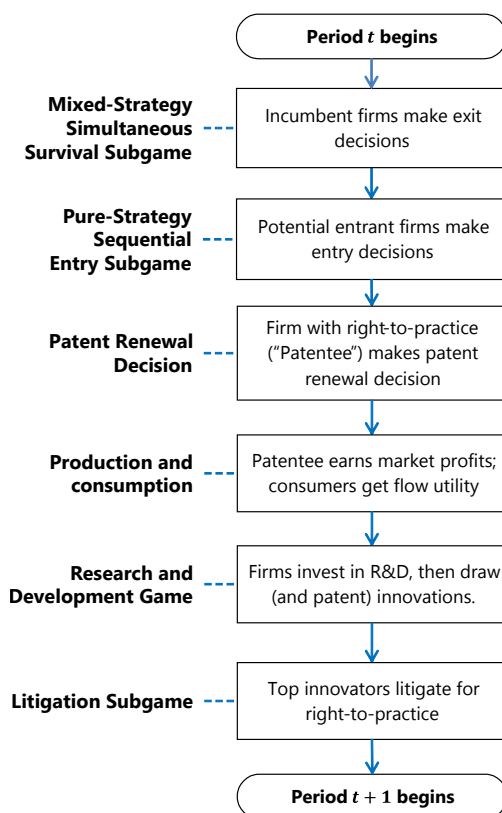


Figure 1.1: Model timing within period

The model is necessarily dynamic so that we can truly evaluate the implications for competition that increasing or decreasing patent strength entail. In a static model, increasing the “strength” of patents mechanically increases competition, as the reward at the end of the “patent race” gets larger. This tends to generate corner solutions in static models: patent duration should be zero and breadth should be infinite, patent duration should be infinity, etc. In a dynamic model, increasing the “strength” of patents has two effects, (1) pro-competitive: increasing the reward for innovation increases R&D efforts, and (2) anti-competitive: intellectual property can discourage innovation by competitors that relies on patented ideas (patent trolling, patent thickets, etc.)

Without dynamics you will only capture the first effect. A two-step model (as in Chang [1995]) will not accurately capture the true value to firms of follow-on innovations. For example, consider fixed costs of entry. When optimally rewarding firms in a static model, higher entry costs would

naturally imply higher patent strength. However, in a dynamic model, higher entry costs increase the value of a patent as they deter future entry (competition). How patent strength should vary in industry primitives then is not a priori known.

The State The state in period t is given by the tuple $\mathbf{s}_t = (N_t, \tau_t, \delta_t, \iota_t)$. $N_t \in \mathbb{N}$ is the number of firms during the research stage of a period. $\tau_t \in \{1, \dots, T\}$ is the statutory number of periods remaining on the current patent. $\delta_t \in \{1, 2, \dots, \Delta\}$ is lagging breadth on the current patent. (That is, any quality within δ_t of the currently patented frontier quality is infringing and cannot be produced.) ι_t is the identity of the patentee. (I use the term patentee as shorthand for “firm with the right to practice the state-of-the-art quality”.) The state space is bounded as N_t is bounded since market profits are bounded and entry/fixed costs are positive. $\mathbf{s}_t$ evolves according to a Markov process that I will discuss below.

1.3.1 Production and Consumption

The output market most closely resembles the single-index quality model found in Grossman and Helpman [1991].³² Quality is denoted by μ_t . To be clear, “quality” denotes consumer quality. Innovation does not mean increasing efficiency of production processes.³³ One can think of the Grossman-Helpmann framework as granting quality-leaders infinite patent duration, with patents immediately voided by subsequent innovation. (Forward protection is zero.) In my model, patents have finite duration. Additionally, recognizing that innovation is often cumulative, patents in my model grant a measure of innovation exclusivity to the patentee. I will make the meaning of “patent” precise in the following sections.

³²Unlike the standard Ericson-Pakes framework (Ericson and Pakes [1995a]), there is no horizontal differentiation in the model. Hopenhayn and Mitchell [2001] examine patent breadth in a dynamic model with horizontal product differentiation, where products vary in their ability to be imitated.

³³This would be more the flavor of Atkeson and Burstein [2011]. Viewing consumers as utilizers of production technology and R&D firms as producers of such technology, this model could apply positively but would have no bearing on true end-user consumers.

R&D firms make period-by-period dynamic decisions on entry, exit, patent renewal, R&D spending and litigation spending to resolve patent disputes. Innovation (steps up the quality ladder) are stochastically determined by the R&D efforts of firms.

Consumers We consider a market of size M_t . Consumer $i \in [0, M_t]$ has unit demand and the utility function over qualities μ and prices $p(\mu)$

$$u_i(\mu) = \lambda_i \mu - p(\mu). \quad (1.1)$$

$\lambda_i \geq 0$ is the “taste-for-quality” and has a continuous cdf F_λ .³⁴ Consumer i chooses μ that maximizes $u_i(\mu)$.³⁵ ³⁶ Although there is no outside good, I allow market size M_t to vary with time as a function of the state: $M_t = M(\mathbf{s}_t)$.³⁷ ³⁸ Consumers and firms discount according to $\beta \in (0, 1)$.

Production Firms and Output Market Equilibrium All quality levels μ are produced at constant marginal cost $\bar{c}$.³⁹ At time t , the frontier quality μ_t might be “patented” with patent breadth δ_t . Production firms are free to produce qualities $\mu \leq \mu_t - \delta_t$. The patentee (the firm who has the right to practice) alone can produce the “state-of-the-art” patented quality μ_t . If there is no active

³⁴I assume $\mathbf{E}[\lambda_i]$ exists.

³⁵Note that the model is one of vertical differentiation. Introducing, for example, logit tastes shocks to (1.1) would introduce horizontal differentiation, but not in any interesting or strategically meaningful way.

³⁶The requirement for stationarity imposes strong restrictions on how quality enters consumer utility to ensure that μ_t does not become a state variable for research firms. One alternative consumer specification more closely follows Grossman and Helpman [1991].

³⁷For example, in a market where the good is considered more durable, I might have $M_t = Mo^v$ where $o \in (0, 1]$ and v is the vintage of the current technology. This doesn’t increase the state space as the firm only cares about vintages $v = T - \tau$.

³⁸Consumers are (unfortunately) captive in the sense that there is no outside good. For some product lines this assumption is more palatable than for others: consider cancer treatment versus HDTVs. Introducing an outside good necessitates taking a stand on its quality process relative to the inside goods. An exogenous outside innovation process implies that the share of the outside good deterministically goes to either zero or one, neither of which are interesting cases. Other assumptions tying the outside innovation rate to the inside innovation rate either make the outside good uninteresting or imply the long run inside innovation rate is the same as the outside innovation rate, which is undesirable in a model of innovation. See Goettler and Gordon [2014] for a discussion of this issue and how it relates to Pakes-McGuire Pakes and McGuire [1994]. An interesting (though potentially intractable) avenue of research for this paper would be a “Dynamic Programming Squared”-type exercise Ljungqvist and Sargent [2004] in which I solve the model repeatedly for several distinct markets, where the quality growth rate of the outside good is the equilibrium quality growth in the other product lines.

³⁹Adding non-constant marginal cost in quantity poses no difficulties. Cost to quality *does* prevent the firms from having a stationary problem, unfortunately.

patent, all firms are free to produce the frontier quality μ_t . I assume static Bertrand pricing in the output market. There is a competitive fringe on entry in the production market. (I will loosely call qualities $\mu < \mu_t$ “generics”, and the state-of-the-art “branded”. Note that these production firms with the exception of the patentee are not the same as the R&D firms that the model studies.)

In the case of no active patent (that is, any relevant patents to the state of the art have expired), all production firms are free to produce at the state-of-the-art quality μ_t . Bertrand competition ensures a price $p(\mu_t) = \bar{c}$. All M_t consumers consume the highest quality and there is zero surplus on the supply side.

In the case of an active patent, the patentee is free to choose a monopolist price $p_t = p(\mu_t)$, and he competes for demand with “generics” $\mu \leq \mu_t - \delta_t$ priced at marginal cost. Given the Bertrand assumption on generics, observe that if a consumer consumes a generic quality, it will be $\mu = \mu_t - \delta$. Consumer i purchases the branded quality only if his taste for quality satisfies $\lambda_i \geq \frac{p_t - \bar{c}}{\delta_t}$. Therefore, the patentee solves the static pricing problem

$$\max_{p_t \geq \bar{c}} \underbrace{M(\mathbf{s}_t)}_{\text{Market Size}} \times \underbrace{(p_t - \bar{c})}_{\text{Markup}} \times \underbrace{\left(1 - F_\lambda\left(\frac{p_t - \bar{c}}{\delta_t}\right)\right)}_{\text{Market share}}.$$

Letting $z = \frac{p_t - \bar{c}}{\delta_t}$, this is equivalent to the problem

$$\max_{z \geq 0} z(1 - F_\lambda(z)). \quad (1.2)$$

Letting z^* denote the solution to (1.2), I have the pricing and profit functions

$$\begin{aligned} p^*(\delta_t) &= \bar{c} + \delta_t z^*, \\ \pi(\mathbf{s}_t) &= M(\mathbf{s}_t) \cdot \delta_t z^* (1 - F_\lambda(z^*)). \end{aligned}$$

Prices and profits scale linearly in δ . Under an active patent,⁴⁰ given the patentee pricing I may write consumer stage surplus from consumption as

$$\begin{aligned}\tilde{CS}_t &= M(\mathbf{s}_t) \cdot \left(\underbrace{\int_0^{z^*} (\lambda_i [\mu_t - \delta_t] - \bar{c}) dF_\lambda(\lambda_i)}_{\text{consumers buying generics}} + \underbrace{\int_{z^*}^{\infty} (\lambda_i \mu_t - [\bar{c} + \delta_t z^*]) dF_\lambda(\lambda_i)}_{\text{consumers buying branded}} \right) \\ &= M(\mathbf{s}_t) \cdot \left(\underbrace{\mathbf{E}[\lambda_i] \mu_t}_{\text{mean utility}} - \underbrace{\delta_t \mathbf{E}[\lambda_i | \lambda_i \leq z^*] F_\lambda(z^*)}_{\text{rationing}} - \underbrace{\delta_t z^* (1 - F_\lambda(z^*)) - \bar{c}}_{\text{monopolist rents}} \right) \quad (1.3)\end{aligned}$$

Through patent policy, society trades rationing and rents for growth in mean utility $\mu_t \mathbf{E}[\lambda_i]$. Our demand system is simple, but not uncommon in the much of the partial-equilibrium patent literature (see O'Donoghue et al. [1998], Hopenhayn and Mitchell [2010], O'Donoghue [1998], Rockett [2010]). Since my goal is to examine the incentives and innovation rather than develop a demand framework, I believe this is defensible.⁴¹

Exogenous Innovation The model allows for exogenous innovation to occur through outside channels such as government or academia. I assume that following the resolution of the patent state and patentee identity, an exogenous quality step $\delta^g \in \{0, \dots, \bar{\delta}\}$ is drawn from the pmf $\eta^g(\delta^g)$. If $\delta^g = 0$, nothing happens, otherwise the existing patent is voided. I generally take $1 - \eta^g(0)$ to be small.

⁴⁰With no active patent, $CS_t = M_t(\mu_t \mathbf{E}[\lambda_i] - \bar{c})$.

⁴¹An interesting modification of my demand system could have naive consumers purchasing durable goods. Consumer i owns quality $\tilde{\mu}_{it}$ when he makes his decision. Consumer i has a two-stage problem:

1. Conditional on upgrading, decide between μ_t and μ_{t-1} (if there is a patent).
2. Conditional on the optimal upgrade, decide whether or not to upgrade.

Utility is then

$$u_{it} | \tilde{\mu}_{it} = \max \{ \lambda_i \tilde{\mu}_{it}, \max \{ \lambda_i \mu_t - p(\mu_t), \lambda_i (\mu_t - \delta) - \bar{c} \} \}.$$

This necessitates the distribution over $\tilde{\mu}_{it}$ as a state variable. Even more interestingly, I could have a truly dynamic consumer problem involving consumer expectations over quality growth as in Goettler and Gordon [2011].

1.3.2 Government

Reflecting the multifaceted nature of what can be termed “patent policy”, I allow “government” in my model a variety of policy instruments. The most straightforward, is statutory patent duration $T \in \mathbb{N}$. Patent duration is set by the Congress. In my model, an active patent entails production exclusivity on the frontier quality and partial research exclusivity for T periods. Note that statutory patent duration does not equal *effective* patent duration, the valuable life of the patent.

Congress may set a patent maintenance schedule, where renewing a patent has cost $a(\mathbf{s}) \geq 0$ where $\mathbf{s}$ is the state.^{42 43 44}

A less well-defined aspect of patent policy is defining *patent breadth*. Following O’Donoghue et al. [1998], I distinguish between *leading* and *lagging* breadth. Lagging breadth denotes the quality distance for lesser qualities from the patented state to be non-infringing. In this model, lagging breadth is simply δ_t .⁴⁵ Leading breadth refers to the minimum size of innovations over the patented state-of-the-art to be non-infringing. This is sometimes called *patentability*: how “innovative” or non-obvious an innovation must be to be rewarded a patent. The model can accommodate this concept, but for simplicity I assume that all innovations are sufficiently innovative to be given patents.⁴⁶ In this model, leading breadth is a function of litigation and is determined both exogenously by court protection for patentees from follow-on innovation, and endogenously through litigation spending.⁴⁷ (The model also allows for Congress to mandate cost subsidies to the firms. $\bar{\alpha}^x, \bar{\alpha}^f, \bar{\alpha}^e \in [0, 1)$ denote *ad valorem* subsidies to R&D expenditure, fixed costs and entry costs, but I don’t study these specifically.)

⁴²See <http://www.uspto.gov/patents/process/maintain.jsp> for the US patent maintenance fee schedule.

⁴³It is of potential interest in future work to allow $a(\mathbf{s})$ to be negative, that is, rewarding innovations directly.

⁴⁴This cost captures the “patents as options” reasoning in Pakes [1986]. In his model, firms make period-by-period decisions about patent renewal, facing a patent maintenance fee schedule. Pakes uses these decision rules to estimate the value of specific patents.

⁴⁵In a more complex model of horizontal differentiation, patent breadth in characteristic space can be thought of as a ball around your patented characteristic bundle into which competitors may not tread. In my vertically differentiated framework, quality is the only distinguishing characteristic.

⁴⁶Given my discrete quality ladder, this seems natural.

⁴⁷Theoretically, both leading and lagging breadth are outcomes of the court system. I only model the court system for leading breadth and do not allow for infringement in production.

1.3.3 R&D Firms

Firms are expected net-present-value profit maximizers with discount rate $\beta \in (0, 1)$. When I say “firm” I mean a research line for a single good/type-of-good within a possibly larger firm. Research lines are atomistic in the sense that each research line acts independently. Entry and exit should be thought of as a firm starting and scrapping, respectively, a research program. Excepting possible distinctions between patentees and non-patentees, firms are taken to be homogeneous with identical productivity. As the model is not being used in the analysis of firm-level microdata, the easing of the computational burden that homogeneity grants compensates for the lack of richness.⁴⁸ Firm homogeneity allows us to conduct detailed policy experiments that would be infeasible in a heterogeneous-firm setting.

Timing Timing is outlined in diagram 1.1. Within a period: (1) incumbent firms decide whether to scrap, (2) entrants decide whether to enter, (3) the patentee decides whether to renew his patent with the patent office, (4) firms earn market profits from production, (5) firms invest in R&D and draw quality innovations and (6) conditional on innovation draws, firms litigate to resolve which firm has the right-to-practice. (The output stage is “offline”, which all other decisions are dynamic controls.) See figure 1.4 for a diagram of stages (1)-(3) and figures 1.2, 1.3 for diagrams of stages (4)-(6), for the cases of an inactive and active patent, respectively.

Value functions and policies Each stage outlined above necessitates the computation of optimal policies and value functions, reflecting the evolution of the state within a period. In all subsequent value functions, the subscript “ ι ” denotes patentee status. $\iota = P$ denotes the patentee, $\iota = 0$ denotes a non-patentee firm. I must solve for

⁴⁸My only concern is that this shuts off any sense of long-term projects. In a time-to-build sense, one could imagine R&D investment not necessarily resulting in immediate innovation, but instead advancing a research program to position the firm with a likelier chance of innovation in subsequent periods. Adding this facet to the dynamic investment decision would be interesting and remains an avenue for future extension. The presence of sunk entry costs this feature in some sense – a 1% increased probability of innovation has different average total cost for an incumbent and entrant.

- $V_{SG}(\iota, N, \tau, \delta)$, the value to an incumbent of playing survival (exit) game with N incumbent firms.
- $V_{EG}(\iota, \tilde{N}, \tau, \delta)$, the value to an incumbent of playing the entry game with $\tilde{N}$ surviving incumbent firms. $V_{EG}^e(\tilde{N}, \tau, \delta)$ denotes the value function for entrants.
- $V_R(\iota, N, \tau, \delta)$, the value of entering the patent-renewal stage, before renewal decisions or patent obsolescence shocks.
- $V(\iota, N, \tau, \delta)$, the value of making it to the R&D stage. (This is the value function of principal interest.)
- $V_\ell(\iota, N, \tau, \delta, \{\delta_i\}_{i=1}^N)$, the value of making it to the litigation stage conditional on the state s and the innovation draws $\{\delta_i\}_{i=1}^N$.

Associated with these value functions are policies on entry, exit, patent-renewal, R&D investment and litigation spending that will be detailed below. I unwrap the period in reverse-order.

1.3.3.1 The Litigation Subgame

Litigation⁴⁹ entails costs to the firms as well as social costs from running the court system. I assume that if there is total litigation spending L , then society incurs a cost $\kappa^\ell L$, $\kappa^\ell > 0$.⁵⁰

Suppose that each firm $i = 1, \dots, N$ has drawn an innovation $\delta_i \in \{0, 1, \dots, \Delta\}$. Take the value function from the survival subgame at the start of the next period, V_{SG} , as given. Regardless of litigation outcome and the assigned lagging patent breadth δ' , I assume that $\mu' = \mu + \max_i \delta_i$.

When there is no active patent, I assume that only the top innovators “win” the patent race and litigate for right to practice. That is, let $\delta^{\max} = \max_i \delta_i$ and let $n^{\max} = \#\{i \mid \delta_i = \delta^{\max}\}$. If

⁴⁹At this point the reader might ask why the model doesn't have licensing. In a complete information model, it is difficult to have both licensing and litigation simultaneously in the model without either firms holding different beliefs, or some sort of dynamic strategic incentive to litigate rather than license (to deter future innovators). Since the current policy debate focuses more on patent trolling and litigation, I have chosen to focus on litigation rather than licensing in this model. An expansion of this model could try to accommodate both modes of patent assignment.

⁵⁰ In recent years, economists and law professors have documented an explosion of patent litigation. There are some correlates with R&D spending, firm size, patent portfolios. Bessen and Meurer [2014] estimates the social cost of patent litigation at \$83bn a year. This is what κ^ℓ is meant to capture.

$\delta^{\max} = 0$, no innovation occurred and I go to the next period. If $\delta^{\max} > 0$ and $n^{\max} = 1$, it is as if one firm has uniquely won the patent race and that firm is assigned a patent $(T, \delta^{\max})$. If, however, $\delta^{\max}, n^{\max} > 0$, multiple firms simultaneously win the patent race, and must litigate to resolve which firm has the right to practice. Let the patent race winners $j = 1, \dots, n^{\max}$ each spend ℓ_j on litigation. With probability $\chi_j = \chi(\ell_j, \ell_{-j})$,⁵¹ firm j is assigned a patent $(T, \delta^{\max})$. Let $\iota_L \in \{0, 1\}$ denote whether a given firm gets to litigate as a function of $\{\delta_i\}_i$. This allows us to define the value function for litigation,

$$V_\ell(\iota_L, N, \delta^{\max}, n^{\max}) = \begin{cases} \beta [\chi_j V_{\text{SG}}(P, N, T, \delta^{\max}) + (1 - \chi_j) V_{\text{SG}}(0, N, T, \delta^{\max})] - \ell_j, & \iota_L = 1 \\ \beta V_{\text{SG}}(0, N, T, \delta^{\max}), & \iota_L = 0 \end{cases}$$

When litigation doesn't occur, I let $\iota_L = 0$ for all and

$$V_\ell(N, \delta^{\max}, n^{\max}) = \beta V_{\text{SG}}(\iota' = \iota ? P : 0, N, \tau', \delta').$$

Firms choose ℓ_j given ℓ_{-j} .

When there is an active patent, litigation is considerably more complicated, as now the patentee has forward exclusion rights on his competitors. I make the following simplifying assumptions.

Assumption 1.1. *If the patentee innovates more than any competitor, the patentee retains the right to practice without any litigation.*⁵²

⁵¹ χ is increasing and concave in its first argument and $\chi(0, \cdot) = 0$.

⁵²The code also allows for one to specify that the top non-patentee innovators and the patentee always go to court, even if the top non-patentee innovators innovate less than the patentee. Qualitatively, the results were unchanged but of course equilibrium quantities did change.

Assumption 1.2. *Only the top non-patentee innovators “win” the patent race. If they innovate more than the patentee, they get to litigate with the patentee over the right to practice next period.*⁵³

54

Let $\delta^C = \max_{i \neq P} \delta_i$ and let $n^C = \{i \neq P \mid \delta_i = \delta^{\max}\}$ denote the maximum innovation of non-patentees and the number of firms who achieved that draw, respectively. Let δ^P denote the innovation draw of the patentee. If $\delta^P = \delta^C = 0$, there was no innovation and the patent clock winds down: $\tau' = \max\{\tau - 1, 0\}$ and $\delta' = \delta$. If $\delta^P > \delta^C$, the patentee builds on his patent according to patentee building rules

$$\tau' = \tau^{\text{NL}}(\mathbf{s}, \delta^P), \quad \delta' = \delta^{\text{NL}}(\mathbf{s}, \delta^P).$$

If $\delta^C \geq \delta^P$, some competitors to the patentee have “won” the patent race but the patentee still has forward research exclusion. He can allege in court that the innovators have infringed on his current patent. Then the firms $j = 1, \dots, n^{\max}, P$ each choose litigation spending ℓ_j given ℓ_{-j} . The right to practice is then awarded probabilistically. Firm j wins the right to practice with probability $\chi_j = \chi(\ell_j, \ell_{-j} \mid \mathbf{1}_{\{j=P\}}, N, \tau, \delta, \delta^C, \delta^P, n^C)$. If a non-patentee challenger wins in court, he is assigned a patent ($\tau = T, \delta = \delta^C$). If the patentee wins, the patentee builds on his patent according to patentee building rules

$$\tau' = \tau^{\text{L}}(\mathbf{s}, \delta^C, \delta^P, n^C), \quad \delta' = \delta^{\text{L}}(\mathbf{s}, \delta^C, \delta^P, n^C).$$

Let χ_P denote the probability that the patentee wins, and χ_j denote the probability that non-patentee litigator j wins. I allow the probabilities to depend on the current state $\mathbf{s}$, the number of challengers n^C as well as the relative magnitudes of the patentee’s and challengers’ innovations (δ^P vs. δ^C).

⁵³This patent reassignment process may seem artificial, and it is certainly an abstraction. I’d point out the 1993 American court case, *Richardson vs. Suzuki Motors*, in which patents were reassigned by the courts. The courts do reassign patents. They also invalidate patents issued by the USPTO (for example, see Williams 2013, who studies the invalidation of the Celera patent on the human genome sequencing). At times, they split baskets of patents between litigating firms (see the assignment of the patents for the BRCA1 and BRCA2 breast cancer genes).

⁵⁴A more appealing specification might have firms making binary decisions over suing/counter-suing and sequential resolution of one-on-one court cases. The paper originally modeled the litigation stage in this manner, but computationally adding this many discrete choice variables make equilibrium calculation too unstable.

Let $\iota_L \in \{P, 0, \text{NL}\}$ denote the identities of the patentee, non-patentee litigators and non-patentee non-litigators. The value functions are then

$$\begin{aligned}
V_\ell(\iota_L = P, N, \tau, \delta, \delta^{\max}, \delta^P, n^{\max}) &= \beta \left[\chi_P V_{\text{SG}}(P, N, \tau^L(\cdot), \delta^L(\cdot)) \right. \\
&\quad \left. + (1 - \chi_P) V_{\text{SG}}(0, N, T, \delta^C) \right] - \ell_P \\
V_\ell(\iota_L = 0, N, \tau, \delta, \delta^{\max}, \delta^P, n^{\max}) &= \beta \left[\chi_P V_{\text{SG}}(0, N, \tau^L(\cdot), \delta^L(\cdot)) + \chi_j V_{\text{SG}}(P, N, T, \delta^C) \right. \\
&\quad \left. + (1 - \chi_P - \chi_j) V_{\text{EG}}(0, N, T, \delta^C) \right] - \ell_j \\
V_\ell(\iota_L = \text{NL}, N, \tau, \delta, \delta^{\max}, \delta^P, n^{\max}) &= \beta \left[\chi_P V_{\text{SG}}(0, N, \tau^L(\cdot), \delta^L(\cdot)) \right. \\
&\quad \left. + (1 - \chi_P) V_{\text{SG}}(0, N, T, \delta^C) \right]
\end{aligned}$$

When litigation does not occur, I let

$$V_\ell(\iota_L, N, \tau, \delta, \delta^P, \delta^{\max}, n^{\max}) = \beta V_{\text{SG}}(\iota' = \iota ? P : 0, N, \tau', \delta').$$

1.3.3.2 Research and Development

Preceding the litigation stage is the R&D stage. Firms pay fixed R&D cost $\bar{x}^f$ here. Firm $i = 1, \dots, N$ chooses an R&D spending amount x_i . If $\bar{a}^x > 0$, the firm only pays $(1 - \bar{a}^x)$ fraction of x_i , with the rest covered by the government. R&D investment is stochastically transformed into innovations. Given $(x_1, \dots, x_N)$, firm i draws a quality step $\delta_i \in \{0, 1, \dots, \Delta\}$ according to the pmf $\eta(\delta_i | \iota, x_i, x_{-i})$. η can allow for both crowd-out and spillovers, and the current patentee may also have an advantage in R&D.

When there is no active patent, the firm chooses x_i given x_{-i} to maximize

$$V(N) = \mathbf{E}[V_\ell(\iota_L, N, n^{\max}, \delta^{\max}) | x_i, x_{-i}] - (1 - \bar{a}^x)x_i - (1 - \bar{a}^f)\bar{x}^f.$$

When there is an active patent, the patentee chooses x_P and the non-patentees $j \neq P$ choose x_i to maximize

$$V(P, N, \tau, \delta) = \mathbf{E} \left[V_\ell (\iota_L, N, \tau, \delta, n^{\max}, \delta^{\max}, \delta^P) \mid x_P, x_{-P} \right] - (1 - \bar{a}^x) x_P - (1 - \bar{a}^f) \bar{x}^f$$

$$V(0, N, \tau, \delta) = \mathbf{E} \left[V_\ell (\iota_L, N, \tau, \delta, n^{\max}, \delta^{\max}, \delta^P) \mid x_j, x_{-j} \right] - (1 - \bar{a}^x) x_j - (1 - \bar{a}^f) \bar{x}^f$$

respectively. Recall that I have defined V_ℓ even for cases where there is no litigation (it is just a convenient placeholder for V_{SG} in this case.)

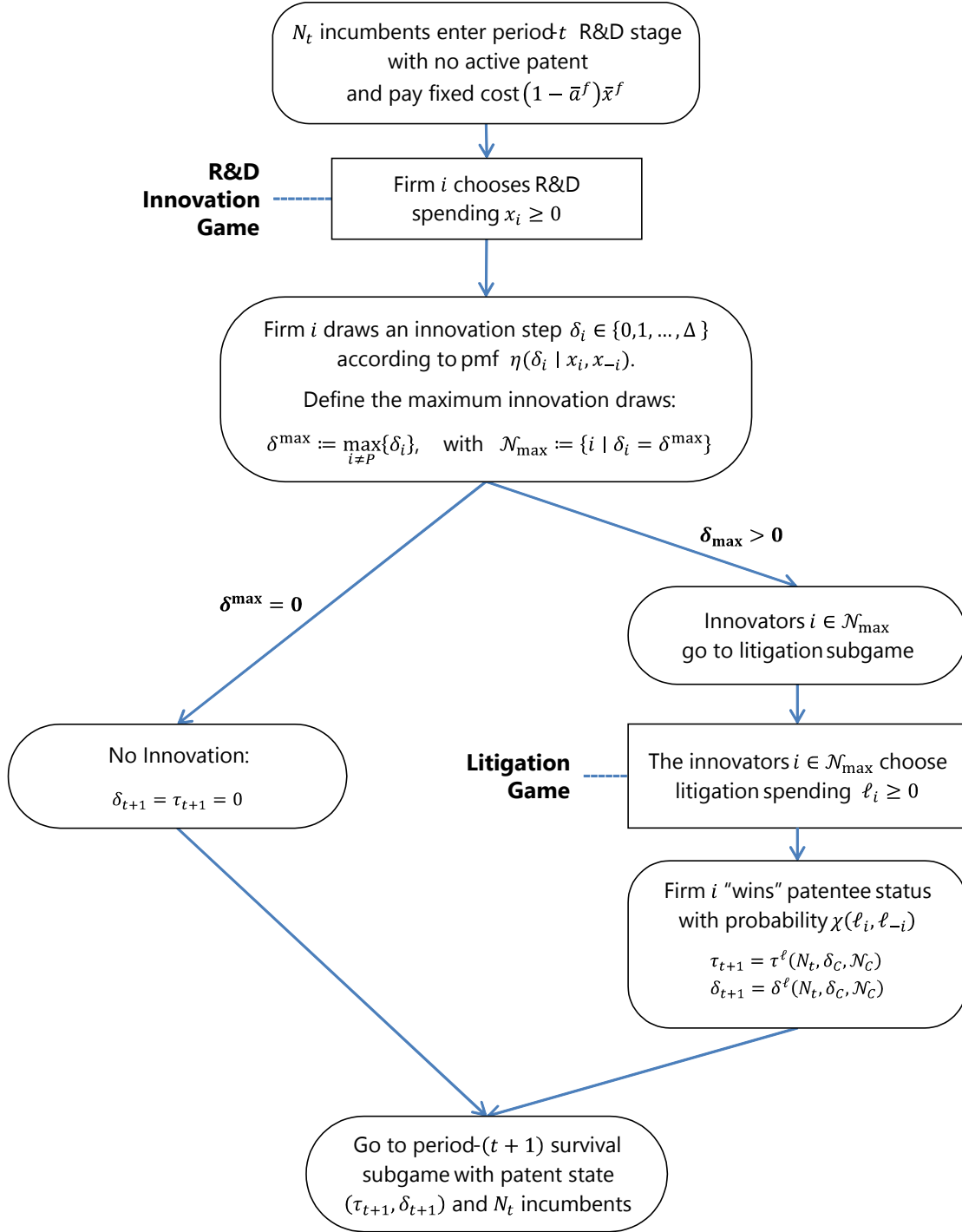


Figure 1.2: R&D and Litigation (No Active Patent)

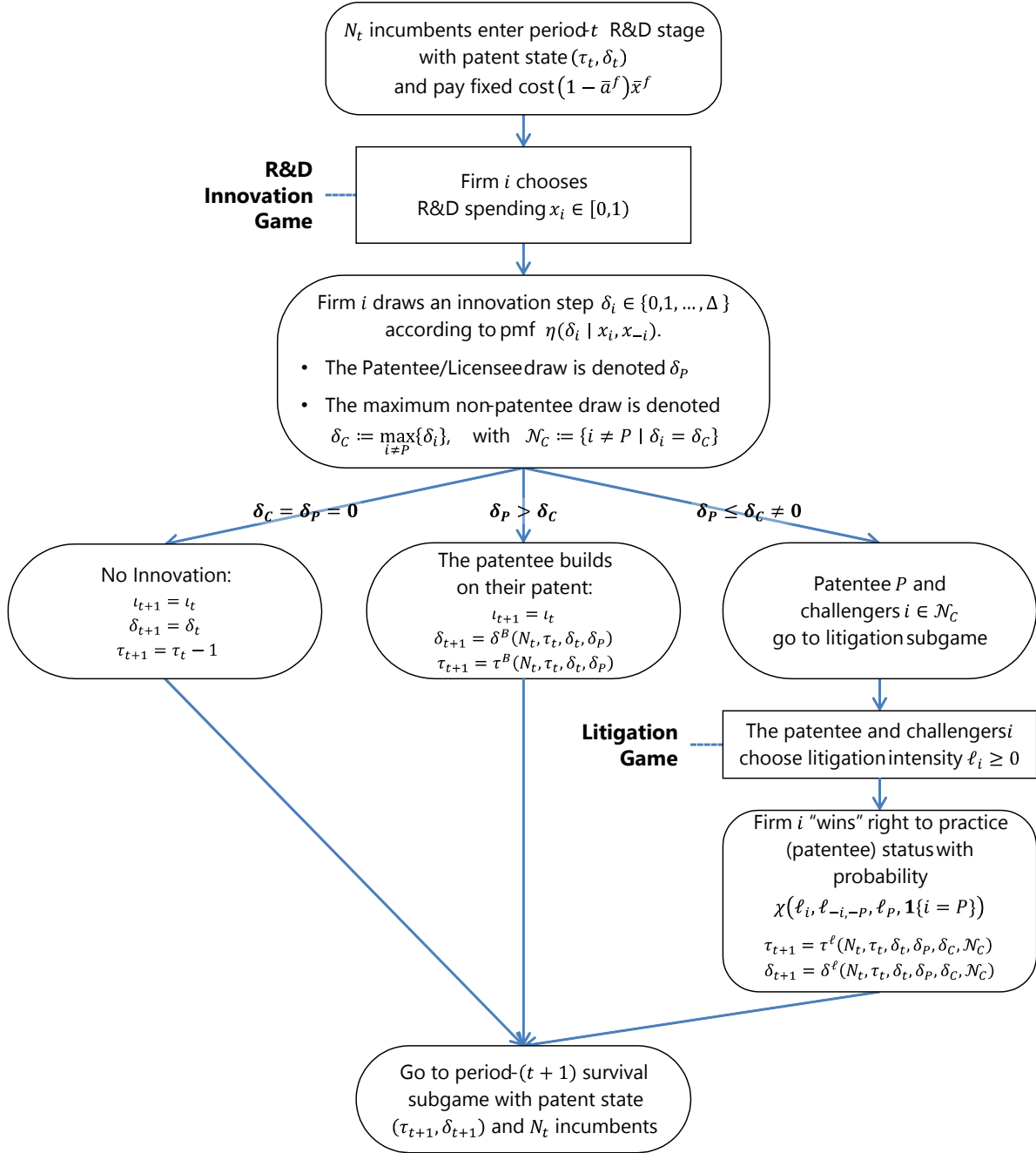


Figure 1.3: R&D and Litigation (Active Patent)

1.3.3.3 Patent Obsolescence and Patent Renewal

With probability $\zeta(s)$, the patent is invalidated. Market tastes may have shifted drastically, or exogenous innovation rendered the patent obsolete, etc. If the patent is not rendered obsolete, I follow in the spirit of Pakes' "Patents as Options" Pakes [1986], and allow patents to be options.

Preceding the R&D stage, the patentee makes decisions about whether or not to renew a patent. In state $\mathbf{s}$ he faces a patent renewal fee $a^R(\mathbf{s}) + \varepsilon^r$. a^R is a function set by policy. ε^r is a small, bounded, mean zero shock that the firm knows when it makes renewal decisions.⁵⁵ It is iid across time and has cdf F_r . The patentee renews if

$$V(P, N, \tau, \delta) + \pi(\mathbf{s}) - [a^R(\mathbf{s}) + \varepsilon^r] \geq V(N),$$

or

$$\varepsilon^r \leq \underline{\varepsilon}^r(\mathbf{s}) := V(P, N, \tau, \delta) + \pi(\mathbf{s}) - a^R(\mathbf{s}) - V(N).$$

Therefore, I can write the value functions for getting to the patent renewal stage as

$$\begin{aligned} V_R(P, N, \tau, \delta) &= (1 - \zeta(\mathbf{s})) F_r(\underline{\varepsilon}^r(\mathbf{s})) \cdot \left(V(P, N, \tau, \delta) \right. \\ &\quad \left. - [a^R(\mathbf{s}) + \mathbf{E}[\varepsilon^r \mid \varepsilon^r \leq \underline{\varepsilon}^r(\mathbf{s})]] + \pi(\mathbf{s}) \right) \\ &\quad + [\zeta(\mathbf{s}) + (1 - \zeta(\mathbf{s})) (1 - F_r(\underline{\varepsilon}^r(\mathbf{s})))] \cdot V(N) \\ V_R(0, N, \tau, \delta) &= (1 - \zeta(\mathbf{s})) F_r(\underline{\varepsilon}^r(\mathbf{s})) \cdot V(0, N, \tau, \delta) \\ &\quad + [\zeta(\mathbf{s}) + (1 - \zeta(\mathbf{s})) (1 - F_r(\underline{\varepsilon}^r(\mathbf{s})))] \cdot V(N) \end{aligned}$$

for patentees and non-patentees, respectively.

1.3.3.4 Entry Subgame

I follow the framework of Abbring et al. [2014].⁵⁶ I model entry sequentially and exit simultaneously. Early versions of the code allowed for both simultaneous and sequential entry/exit models.

⁵⁵ ε^r is a computational necessity more than a deliberate modeling choice.

⁵⁶ Doraszelski and Pakes [2007] write that “Entry has been treated differently in different papers. This proliferation of entry models is largely because there is no agreement, and very little in the way of empirical guidance, on the appropriate way to model entry.”

There did not appear to be a large difference in terms of value functions or welfare for one model versus the other. This may be due to firm homogeneity.⁵⁷

My entry model is sequential with an unbounded number of potential entrants. It is therefore a model of free entry. A potential entrant faces an entry cost $\bar{x}^e + \varepsilon^e$. $\bar{x}^e$ is fixed, and ε^e is a small, bounded market-wide, mean-zero entry cost shock that is iid over time and uniform across entrants. It has cdf F_ε . If the government subsidizes entry ($\bar{a}^e > 0$), the entrant faces the cost $(1 - \bar{a}^e)\bar{x}^e + \varepsilon^e$. Let $\tilde{N}$ denote the number of survivors of the survival game. Conditional on n^e prior entrants, the marginal entrant enters if

$$V_R(0, \tilde{N} + n^e + 1, \tau, \delta) - [(1 - \bar{a}^e)\bar{x}^e + \varepsilon^e] \geq 0$$

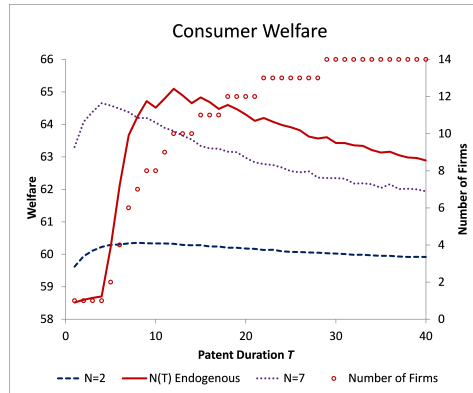
or

$$\varepsilon^e \leq V_R(0, \tilde{N} + n^e + 1, \tau, \delta) - (1 - \bar{a}^e)\bar{x}^e.$$

Let $\bar{\varepsilon}^e$ denote the highest possible entry shock. The maximum number of guaranteed entrants is then

$$n^e(\tilde{N}, \tau, \delta) = \max_n \{ V_R(0, \tilde{N} + n, \tau, \delta) - [(1 - \bar{a}^e)\bar{x}^e + \bar{\varepsilon}^e] \geq 0 \}$$

⁵⁷One might reasonably question the decision to include entry and exit decisions in a homogeneous model. I can shut down entry and exit and instead fix the number of firms at the outset for all t . However, I find that allowing N changes the optimal policy recommendation on duration, for example. In the following figure I compare welfare for N fixed at 2, N fixed at 7 and N free to vary. When N is fixed there does not appear to be much different in the optimal duration, whereas it shifts dramatically to the right when N is allowed to vary.



with a marginal $n^e(\tilde{N}, \tau, \delta) + 1$ entrant with probability

$$p_e = F_e(V_R(0, \tilde{N} + n^e(\tilde{N}, \tau, \delta) + 1, \tau, \delta) - (1 - \bar{a}^e)\bar{x}^e).$$

We can then write the value functions for the entry game⁵⁸

$$V_{EG}(\iota, \tilde{N}, \tau, \delta) = p_e V_R(\iota, \tilde{N} + n^e(\cdot) + 1, \tau, \delta) + (1 - p_e) V_R(\iota, \tilde{N} + n^e(\cdot), \tau, \delta).$$

The value function for the n^{th} entrant is

$$V_{EG}^e(\tilde{N}, \tau, \delta, k, \varepsilon^e) = \max\{V_R(\iota, \tilde{N} + k + 1, \tau, \delta) - [(1 - \bar{a}^e)\bar{x}^e + \varepsilon^e], 0\}.$$

Firm concentration evolves according to⁵⁹

$$N' = N'(\varepsilon^e) = \tilde{N} + n^e(\cdot) + \mathbf{1}_{\{\varepsilon^e \leq V_R(0, \tilde{N} + n^e(\cdot) + 1, \tau, \delta) - (1 - \bar{a}^e)\bar{x}^e\}}.$$

Note that conditional on entry, an entrant and incumbent are indistinguishable. And entrant can play the R&D game immediately upon paying fixed and entry costs.

1.3.3.5 Survival (or Exit) Subgame

At the beginning of a period, incumbent firms make decisions about exit. There is a fixed scrap value $\bar{x}^s$. I assume that firms make simultaneous scrap decisions. Let N denote the current number of firms and $\tilde{N}$ the number of firms after exit.

If $V_{EG}(0, N, \tau, \delta) > \bar{x}^s$, there is no exit and $\tilde{N} = N$. If $V_{EG}(0, 2, \tau, \delta) < \bar{x}^s$, not even one non-patentee firm can exist profitably, and there is complete exit of non-patentees $\tilde{N} = 1$ (the patentee).

⁵⁸I only show the calculation for the case of an active patent; it is analogous for an inactive patent.

⁵⁹If $V_R(0, \dots)$ is weakly decreasing in N , then the above procedures are fairly straightforward. However, the possible presence of cost crowd-out or R&D spillovers (and the nature of the litigation process) means I cannot rule out non-monotonicities in the value functions. In fact, non-monotonicities in the value function of the patentee are common. However, in my experience value functions *at most* have a single local maximum with respect to N – the code works under this assumption. For more exotic cases, the code may break. I have never encountered such cases.

Otherwise, non-patentees can exist profitably in the current state, just not at the current concentration. In this case I solve for the symmetric mixed-strategy Nash equilibrium in exit. The exit probability $p^s(N, \tau, \delta)$ will satisfy the indifference condition

$$\mathbf{E}[V_{\text{EG}}(0, N - n^e, \tau, \delta)] = \bar{x}^s, \quad n^e \sim \text{Binom}(N - 2, p^s(\cdot)).$$

(The $N - 2$ other firms exit with probability $p^s(\cdot)$ leaving the remaining non-patentee indifferent between staying and scrapping, so he is willing to exit with probability $p^s(\cdot)$.) Then $\tilde{N} = 1 + \text{Binom}(N - 1, 1 - p^s(\cdot))$ ⁶⁰ The value functions are then

$$V_{\text{SG}}(\iota, N, \tau, \delta) = \mathbf{E}[V_{\text{EG}}(\iota, \tilde{N}, \tau, \delta)]$$

where $\tilde{N}$ evolves as above. The case for there not being an active patent is analogous and is omitted.

⁶⁰The one complication here is the case of duopoly under an active patent. (A monopolist when there is no active patent never exits or the entire model is broken.) In this case, a single non-patentee is making a scrap decision. A mixed strategy cannot make sense for this firm, so I introduce a tiny scrap value shock ε^s in this case to smooth out the exit decision in duopoly under a patent.

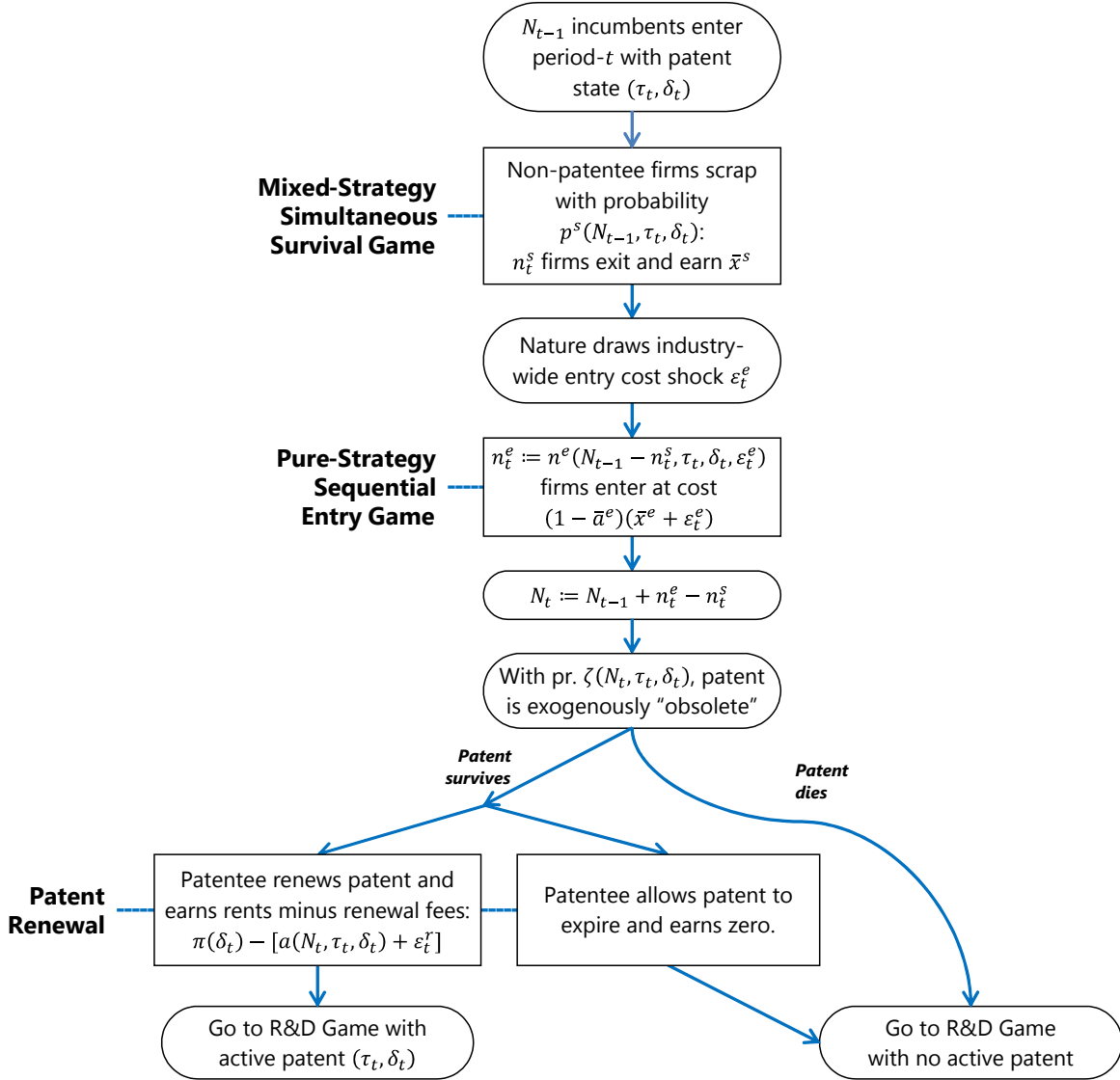


Figure 1.4: Exit, Entry and Patent Renewal

1.3.4 Social Welfare

The component of consumer welfare from consumption decisions, $\tilde{C}S_t$, was given in equation (1.3). Assuming that patent renewal fees go to consumers, and cost subsidization policies and social costs of litigation are borne by consumers, I arrive at the full specification for per-period

consumer welfare,

$$CS_t = \tilde{CS}_t - \underbrace{\bar{a}^e n_t^e \bar{x}^e}_{\text{entry subsidization}} - \underbrace{\left(\bar{a}^x \sum_i x_i + \bar{a}^f N_t \bar{x}^f \right)}_{\text{R\&D cost subsidization}} - \underbrace{\kappa^\ell \sum_i \ell_i}_{\text{social cost of litigation}} + \underbrace{a(s_t)}_{\text{patent renewal fee}}.$$

My welfare measure is then

$$W = \mathbf{E} \left[\sum_{t=0}^{\infty} \beta^t CS_t \right].$$

I do not include any sort of producer surplus in my welfare measure given that in expectation firms earn approximately zero.

1.3.5 Equilibrium

I use the standard conception of equilibrium in the Ericson and Pakes [1995a] flavored industry dynamics literature, symmetric Markov-perfect equilibrium (MPE) Maskin and Tirole [1988, 2001]. A symmetric MPE is one in which all players use the same Markov strategies, conditioned solely on payoff-relevant state variables. In my case, the payoff-relevant variables are N, τ, δ, ι and in the patent litigation subgames also $\delta^{\max}, n^{\max}$ and δ^P, δ^C, n^C for no-active-patent and active-patent, respectively. If I can construct the requisite value functions detailed above such that they are consistent with firm optimization at every stage via policies investment policies $x^*(s)$, and litigation policies $\ell^*(s)$, patent renewal decisions $r^*(s)$, and entry and exit, then I have found a symmetric MPE.⁶¹

⁶¹I am still working on a proof of equilibrium existence using the methods Denardo [1967], Whitt [1980], Doraszelski and Satterthwaite [2007]. The multiple dynamic controls make extending the Doraszelski and Satterthwaite [2007] proofs difficult. Luckily, my algorithm has never failed to converge to what can be verified ex-post as an MPE, and I have never converged to multiple equilibria from different starting values. (I realize that this is a dangerous statement in this literature.)

1.4 Computation and Comparative Statics

We compute symmetric Markov-perfect equilibria using standard Gaussian Pakes and McGuire [1994] value-function iteration methods.⁶² I have not yet encountered a specification for which the code did not converge to an equilibrium⁶³. Once the algorithm gets close to the right amount of entrants and exiters in each state, value functions and policies appear to converge linearly. The value function update operators are not generally contractions. Firm homogeneity (excepting the patentee/non-patentee distinction) ensure that solution time is small. For models that imply a high degree of concentration and only a handful of firms ever being active, the model can converge in a matter of seconds. For models with more competitive outcomes, the algorithm typically takes several minutes.⁶⁴ Computation time grows exponentially in the maximum number of active firms due to the increasing complexity of the binomial distributions which must be frequently calculated for evaluating R&D outcomes, and the quadratic state space in the litigation subgames.⁶⁵

As noted in the online appendix to Doraszelski and Satterthwaite [2007], the choice of iteration scheme alone can produce equilibrium multiplicity in Ericson and Pakes [1995a] styled models. Varying the computational method in my model leads to similar multiplicity in my model.⁶⁶ However, all computed equilibria for a given specification have been relatively close to one another (and are indistinguishable to the naked eye). Borkovsky et al. [2012] notes that in industry dynamics models, equilibrium multiplicity often arises from policy differences in seldom visited states and observationally the sample paths generated by multiple equilibria are comparable. That was for a

⁶²The code will be available on my website, www.princeton.edu/~rkatzwer/optimalPatentsCode.zip.

⁶³The convergence path is often initially rocky until entry and exit have settled down.

⁶⁴The convergence criterion is that relative and absolute error between successive iterates are below $1e-6$ and $1e-3$, respectively and the relative and absolute error between the k th iterate and $(k - 20)$ th iterate are below $1e-5$ and $1e-3$, respectively. All computations were performed on an Intel Core i7-3770 running at 3.4-3.9 GHz. Comparative statics are computed by simulating the MPEs for 400 periods, 15,000 times.

⁶⁵The addition of firm-level heterogeneity would seem to be necessary to bring this model to data. In such a case, computation time would likely require either the stochastic methods of Pakes and McGuire [2001], or translation of the model to continuous time as in Doraszelski and Judd [2012].

⁶⁶For example, varying the frequency of R&D policy update from every ten iterations to every twenty iterations in one specification led to a difference in the expected number of entrants in the state $N = \tau = \delta = 0$, 19.137 vs. 19.066. I keep the update scheme the same across all solutions, and to be safe, update all policies at every iteration (with dampening).

different industry dynamics model, but the same phenomenon appears to hold in the simulations that I ran.

In this section, I will first examine the dynamics of a typical equilibrium. Then I will show welfare comparisons and optimal policy design across a range of environments.

1.4.1 Example Industry Dynamics

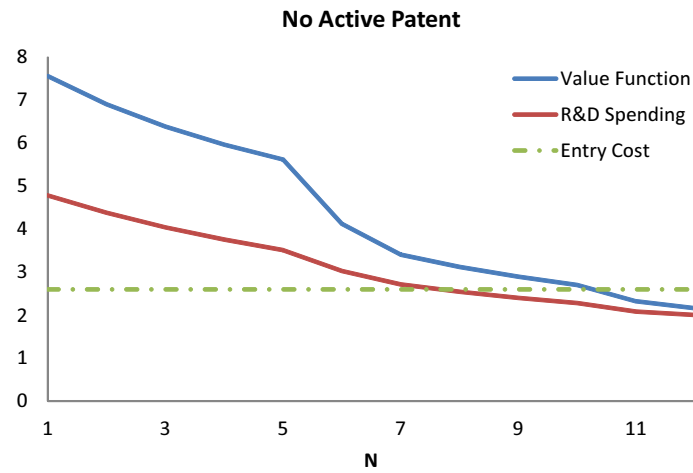
This section will examine a two equilibria of this model. First, we get a sense of firm behavior by examining equilibrium value functions and policies in two different cost settings. I simulate one sample path of the equilibria to give a flavor of the industry dynamics.

I compare two equilibria from settings with identical government policies and primitives. Statutory duration T is 12 and $\Delta = 3$. (The parameters and functional specification can be found in Appendix A.2.) However, in the second setting the fixed costs of R&D increase by 50% and the probability of successful innovation conditional on R&D spending is reduced by 33%.

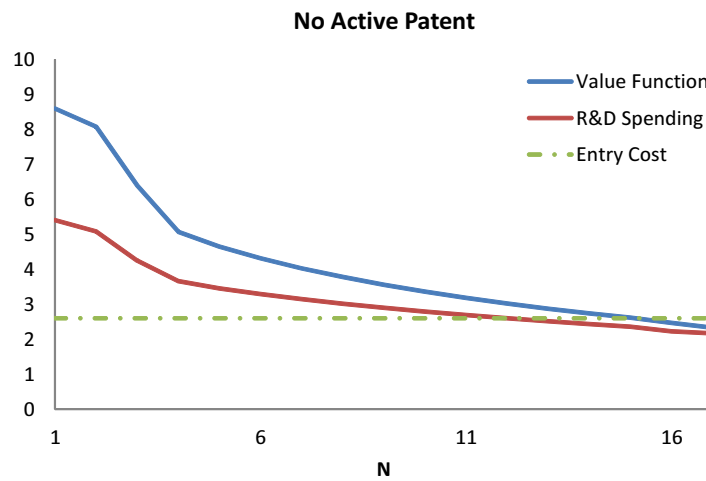
I begin by examining the value functions and investment policies when there is no active patent in figures 1.5a and 1.5b. As one would expect, the value function is decreasing in the number of firms. In the low-cost setting, I have a maximum number of entrants of 10, whereas in the high-cost setting this jumps to 15.⁶⁷ This may seem counterintuitive, but it demonstrates the dynamics of the model nicely: with higher costs, innovation is more costly, and a patentee faces less competition conditional on owning a patent and the forward exclusion that it entails. Therefore, when there is no active patent, R&D firms are ex-ante more valuable. (The non-smoothness in $V_0(N)$ is not an error. For example, in the low cost setting, up to five firms will enter when there is an active patent, this is what generates the kink at $N = 5$.)

Next, I examine behavior when there is an active patent. Figures 1.6 and 1.7 report value functions for patentees and non-patentees in the low-cost setting, respectively. (These are plotted for $N \in \{5, \dots, 10\}$ as these are the number of firms that can occur in these (τ, δ) states in equilibrium.)

⁶⁷The maximum number of entrants when $N = \tau = \delta = 0$ corresponds to the maximum number of firms that can be active in any given period.



(a) Value Function and Investment with no active patent (low cost setting)



(b) Value Function and Investment with no active patent (high cost setting)

Figure 1.5

The value functions for non-patentees are sharply decreasing in the number of competitors. They are very similar across different values of δ , as δ only indirectly effects their future value and innovation rates through the investment choices of the patentee. Any patent that they are potentially awarded is a function of their own innovation and the innovation of the patentee, but not the magnitude of past innovation or current lagging breadth. I see the value functions for non-patentees increase as the patent gets closer to expiry ($\tau \rightarrow 0$). Of course, the opposite holds for the patentee's value functions. Their value is a combination of (i) expected market profits from the right-to-practice (ii) expected value of innovation and (iii) expected value of potentially appropriating the innovations of competitors via infringement lawsuits. (i) generates the vast difference in levels between the patentees value functions across different lagging breadth values δ . (iii) explains why the value function of the patentee decreases less steeply in N than does the value function of non-patentees. In fact, we see that when $\tau = 1$ or 2 (that is the patent is close to expiry), the value function of the patentee is actually *increasing* in N . As the current patent gets close to expiry, the patentee wants more innovators producing (probabilistically) higher-quality innovations that it can then appropriate in court. The patentee uses these appropriated innovations to replace its current, soon-expiring, right-to-practice patent. (The non-patentee value functions “bottoming out” corresponds to the discounted scrap value, $\beta\bar{x}^s = .576$).

Figures 1.8 and 1.9 report R&D investment policies for patentees and non-patentees in the low-cost setting, respectively. Schumpeterian effects dominate the investment decisions of the patentee. Particularly when the patentee has high value patents ($\delta = 2, 3$, and τ close to T), the patentee has no desire to innovate and displace the current patent. The patentee invests very little in R&D. It is more cost effective for him to challenge competing innovators in court than to innovate himself. When τ is close to T , non-patentees invest more in R&D on average than the patentee does. As $\tau \rightarrow 0$, the patentee ramps up investment. The opposite holds for non-patentees: they invest slightly less in R&D as $\tau \rightarrow 0$ on average, since the patentee ramps up (and the advantage

greater investment draws give the patentee in court outcomes⁶⁸) greatly reduces non-patentees' expected return on R&D spending.

⁶⁸Litigation outcomes are functions of litigation spending, and under an active patent, the return to patentee litigation spending decreases in the size of the gap $\delta^C - \delta^P$.

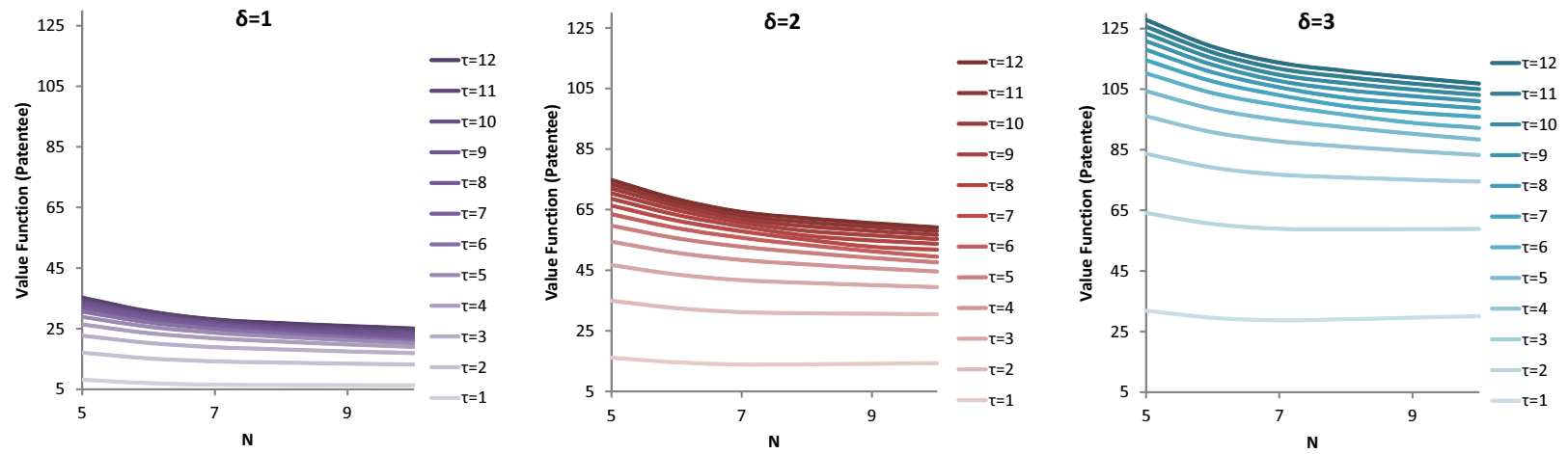


Figure 1.6: Value Functions for Patentee (low cost setting)

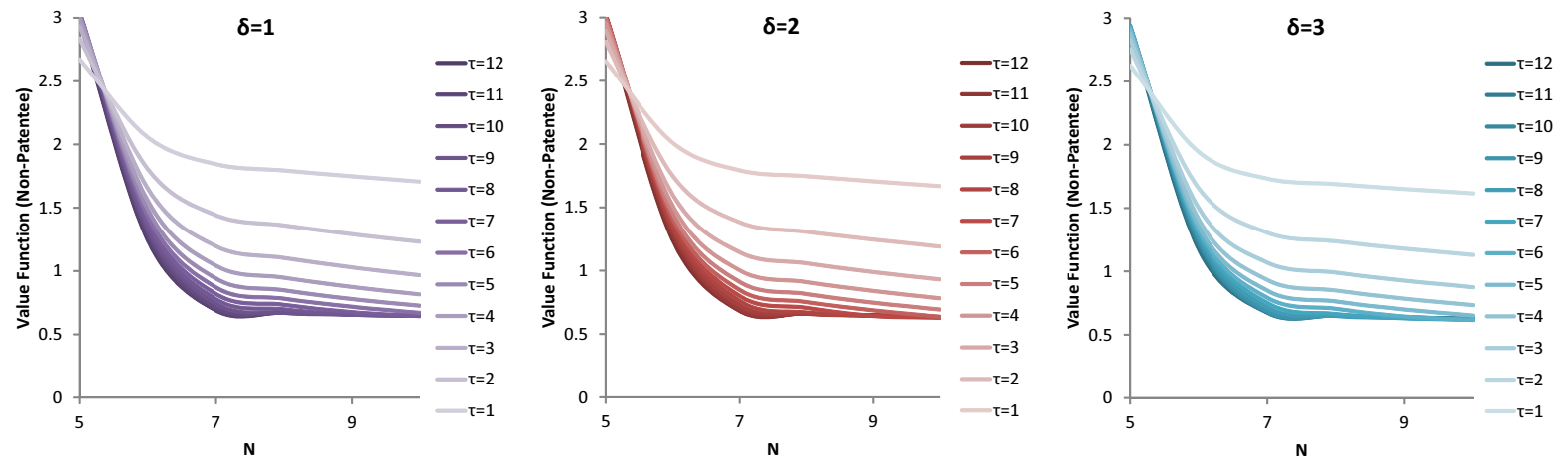


Figure 1.7: Value Functions for Non-Patentee (low cost setting)

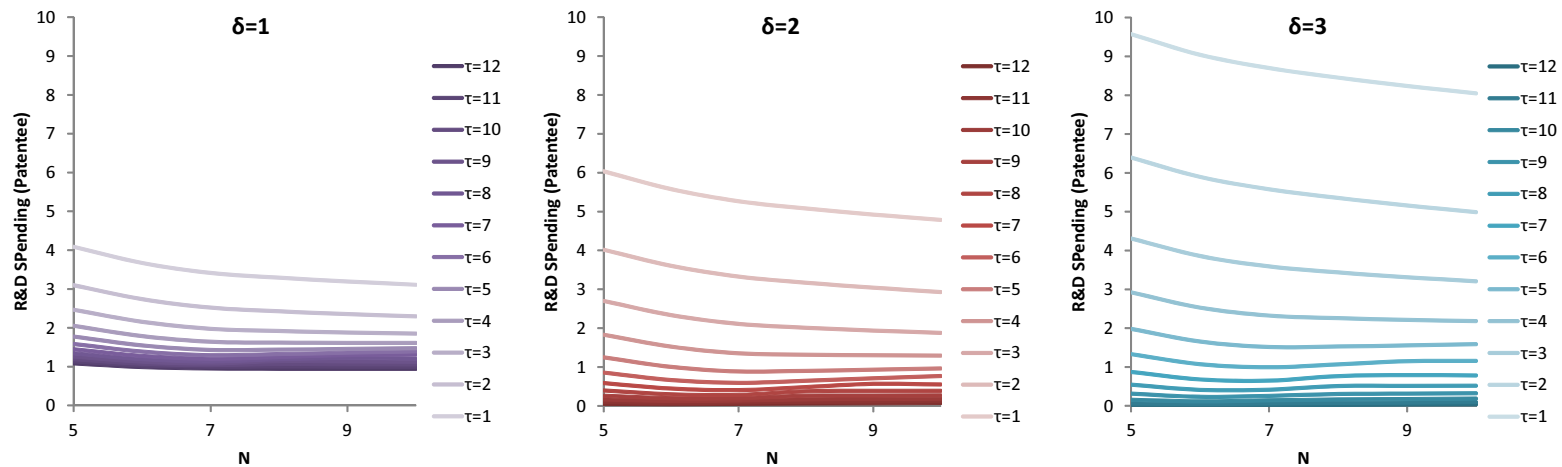


Figure 1.8: R&D Investment for Patentee (low cost setting)

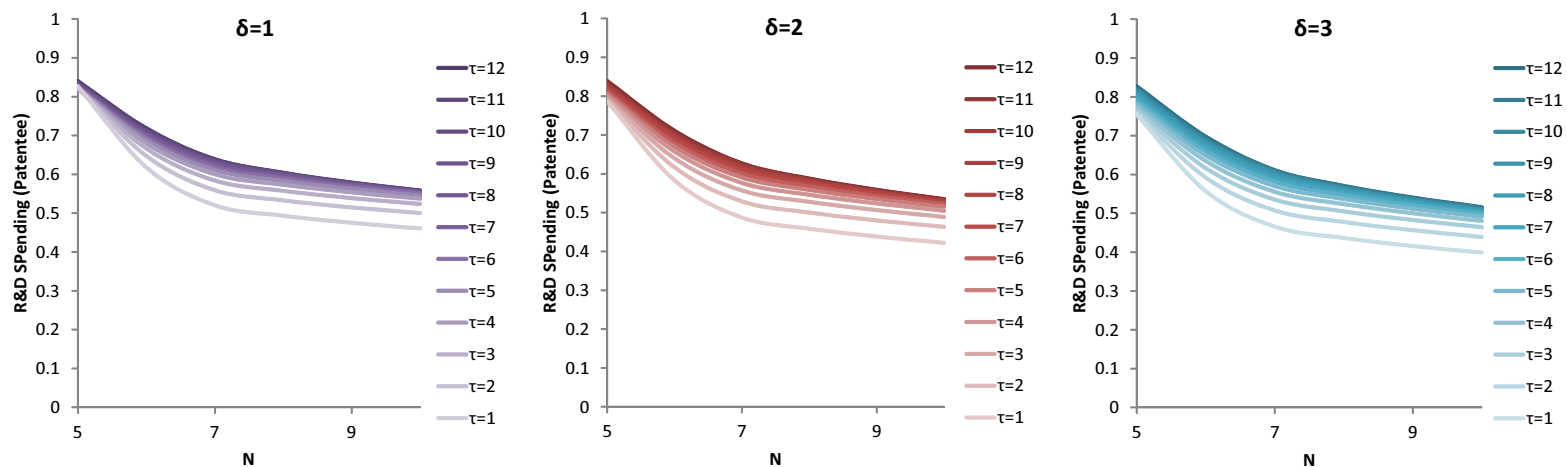


Figure 1.9: R&D Investment for Non-Patentee (low cost setting)

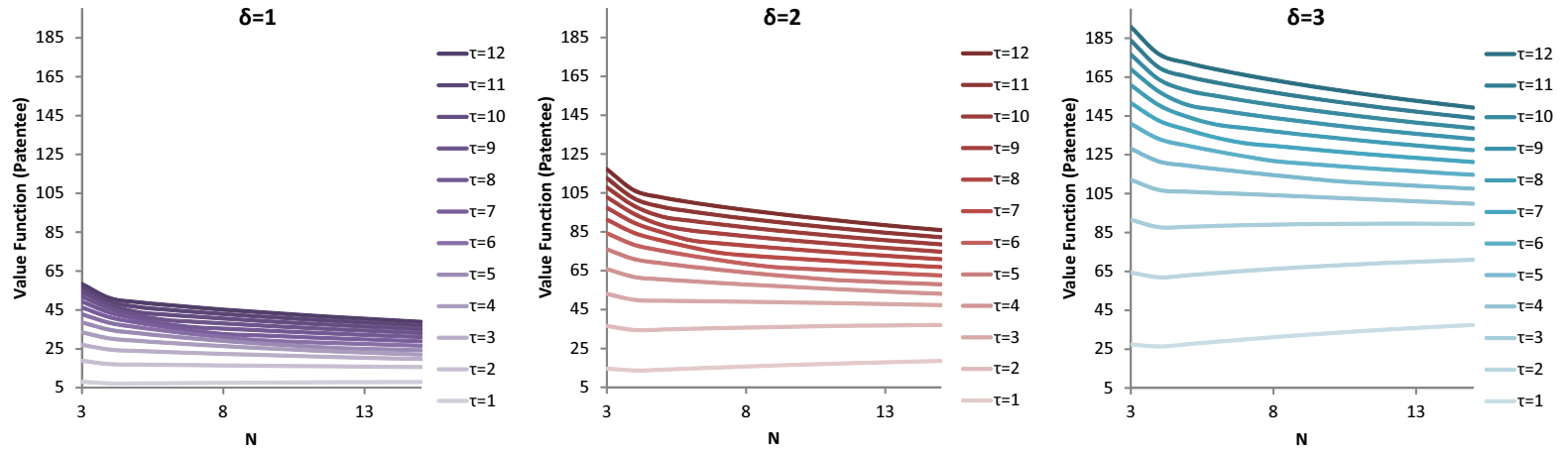


Figure 1.10: Value Functions for Patentee (high cost setting)

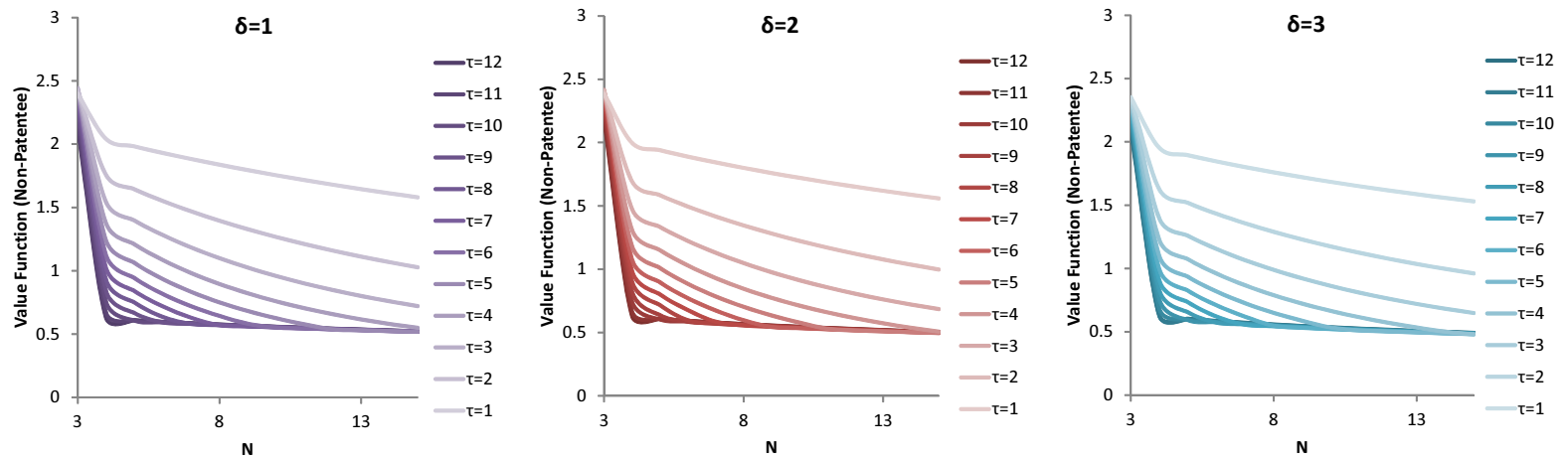


Figure 1.11: Value Functions for Non-Patentee (high cost setting)

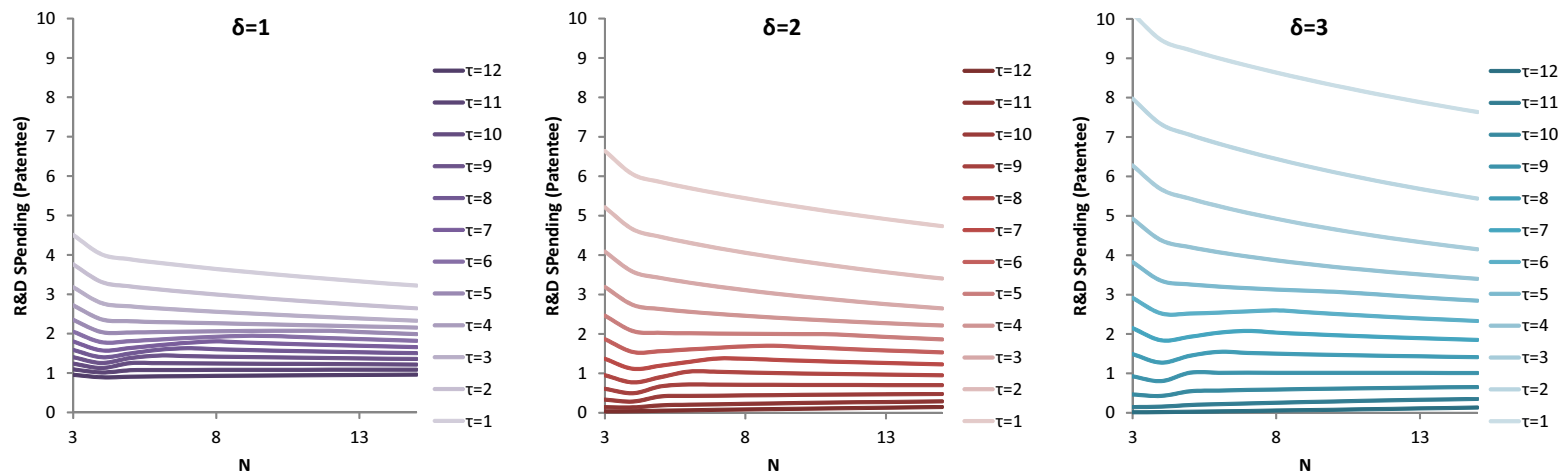


Figure 1.12: R&D Investment for Patentee (high cost setting)

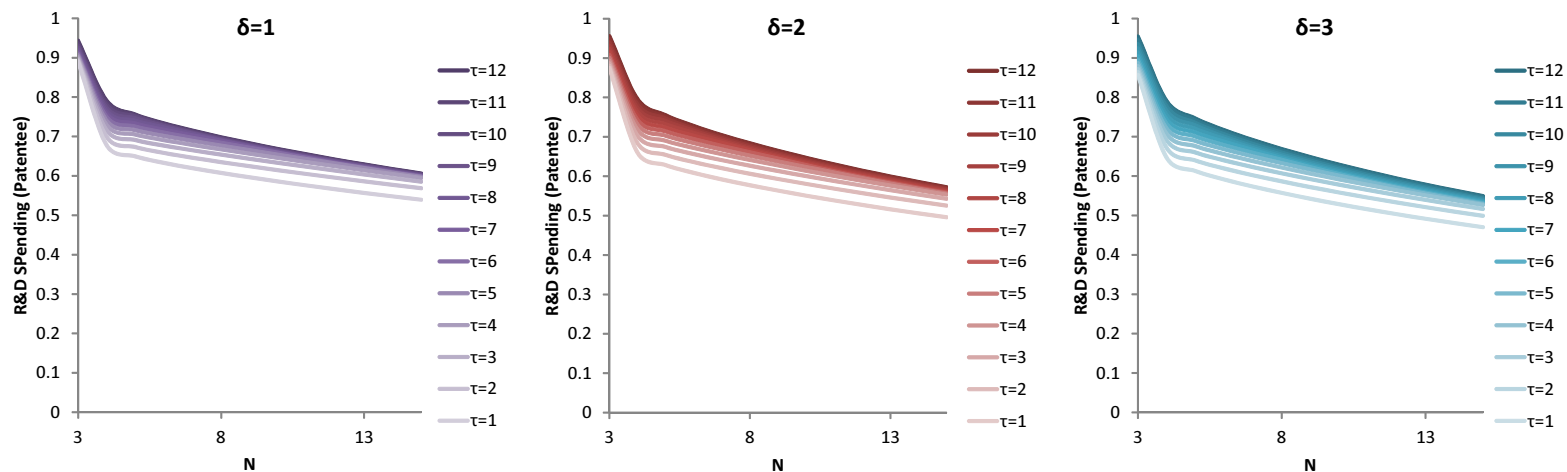


Figure 1.13: R&D Investment for Non-Patentee (high cost setting)

I now turn to the high-cost setting. The equilibrium properties are qualitatively similar to the low-cost setting, but some of the dynamic competition effects are exacerbated. Figures 1.10 and 1.11 report the value functions for patentees and non-patentees under an active patent, respectively. The value functions for non-patentees are very similar to the low-cost setting, but lower, reflecting the more difficult innovation environment non-patentee R&D firms face. The value functions of the patentee, however, are much larger. Higher costs of innovation greatly increase the value of owning a patent. Notice that the slope of the value functions for patentees in N is decreased in the high-cost setting. Competitors are less successful and therefore less of a threat. I see the value function slope upward in N for the patentee for $\delta = 2, 3$ and $\tau = 1, 2, 3$. The value function for the patentees decrease more uniformly in τ in the higher-cost setting. This makes sense: innovation is more costly and occurs less frequently, implying the decrease in the number of market-profitable years remaining has more influence on the value function than in the low cost setting. (Mechanically, patentees and non-patentees spend more on R&D on average because innovation by assumption is more costly here.)

Finally, I compare sample paths of sixty periods of the Markov-perfect equilibrium for the low- and high-cost settings in figures 1.14a and 1.14b, respectively. The equilibrium dynamics are different. In the low cost setting, the entry-and-exit is more muted. N varies between six and ten. Compare this to the high cost setting which oscillates between fifteen firms when there is no active patent and (usually) three under an active patent. (In fact, the market reduces to duopoly in later periods of this sample run.) This reflects the earlier discussion of patents being more valuable ex-ante in the high cost setting, but the environment being more difficult for non-patentees under an active patent. On average, I expect the rate of innovation to be higher in the low-cost setting, and that happens in this sample run. Even though R&D costs are higher in the high-cost setting, *a priori* it isn't certain that R&D spending would also be higher. In fact, in most periods, more is spent on R&D in the low-cost framework. Firms on average spend less in this setting, but more firms exist in equilibrium in most periods. In both settings, I see R&D spending at its highest when there is no active patent (and no patentee waiting in the wings to claim infringement), and R&D

spending in aggregate ramps up as patents approach expiry. Notice the litigation spending spikes. Litigation spending is far higher in the high-cost setting as the reward to winning a court case (and the right-to-practice) is much higher.

1.4.2 Comparative Statics

Having given a flavor of equilibrium dynamics, I now turn to comparative statics on government patent policy.

1.4.2.1 Static vs. Dynamic and Optimal Patent Duration

I first compare optimal statutory duration policy T in static⁶⁹ and dynamic setting. In this section I will examine how optimal duration T^* varies in different environments, but will hold the other patent policies fixed. The environments are indexed by k with lower k indicated higher costs of innovation (or lower probabilities of successful innovation conditional on innovation).⁷⁰ That is, the seven specifications range from very difficult innovation to relative easier innovation. The static model generates the expected welfare and optimal patent duration T^* depicted in figure 1.15. The standard logic holds: more expensive innovations are incentivized by larger rewards (T).

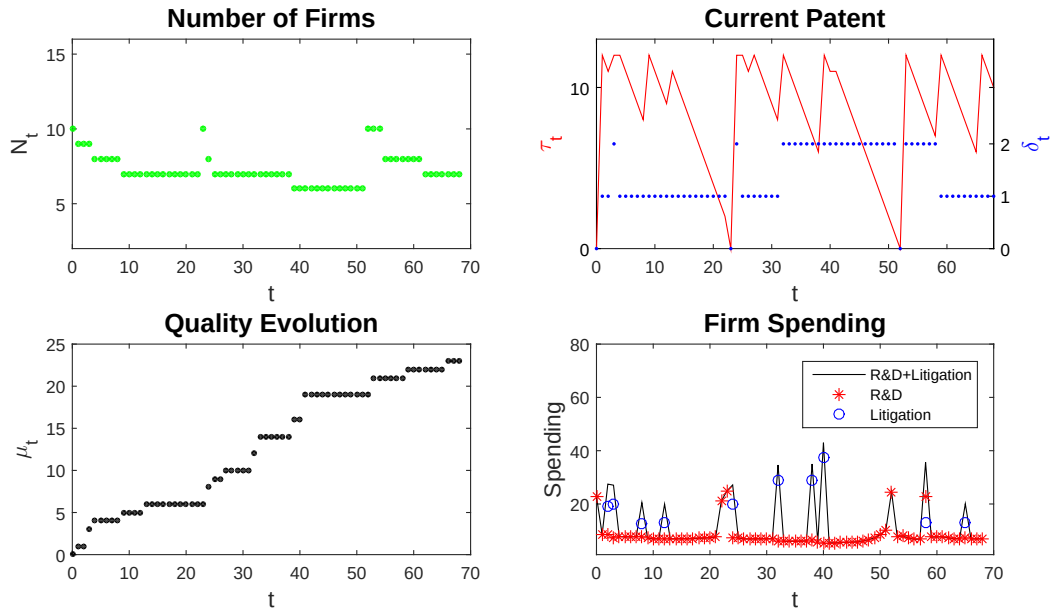
Now, I can repeat the same exercise in the dynamic model. *A priori*, it is not obvious what will happen to T^* . Forward exclusion (which is not captured in a static model) increases the value of holding a patent. This would lead us to believe that the efficient T^* would be *lower* in a dynamic vs. static model. Additionally, while pro-competitive in patent-races with no active

⁶⁹See A.1 for an outline of the baseline static model.

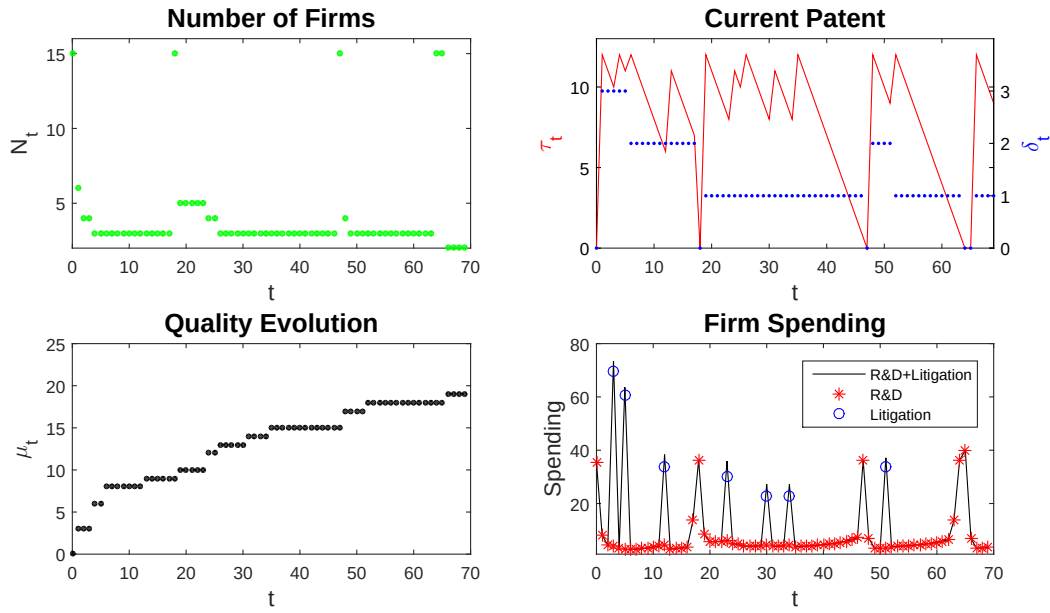
⁷⁰Foreshadowing section (1.4.2.2), I keep the probability (conditional on innovations and litigation) that a current-patentee wins a patent litigation case at

$$\chi_P = \tilde{\chi} + (1 - \tilde{\chi}) \frac{(1 - .1(\delta^C - \delta^P)) \ell_P^{4/5}}{\sum_{j \neq P} \ell_j^{4/5} + (1 - .1(\delta^C - \delta^P)) \ell_P^{4/5}}, \quad \tilde{\chi} = .15. \quad (1.4)$$

That is, a patentee wins a litigation case with at least 10% probability here even if he spends nothing on litigation. $\tilde{\chi} = 1$ is a corner case where the patentee has infinite forward exclusion. In subsequent analysis, I will vary T as well as $\tilde{\chi}$. Consider the same specification as in section 1.4.1. I run seven cost specifications, indexed by k . I let fixed R&D costs vary over $\bar{x}_{(k)}^f = \{.44, .40, .36, .32, .28, .26, .24\}$. Concurrently, the probability of a successful innovation draw conditional on x_i (that is, $\delta_i \neq 0$) is given by $\frac{\alpha_{(k)} x_i^{4/5}}{5 + x_i^{4/5}}$ where $\alpha_{(k)} = \{.12, .14, .16, .2, .24, .28, .32\}$.



(a) A sample MPE path (low cost setting)



(b) A sample MPE path (high cost setting)

Figure 1.14

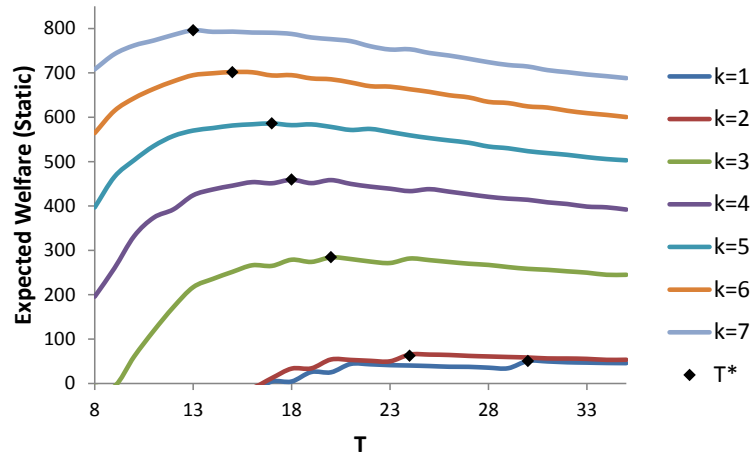


Figure 1.15: Static Welfare as function of T^*

patent, the forward exclusion becomes strongly anti-competitive when there is an active patentee which would also imply that T^* should be lower in the dynamic model to increase competition. However, the dynamic model includes cumulative innovation and the possibility that a patentee's effective patent life is far shorter than the statutory T , which might lead us to expect that T^* is *higher* in the dynamic model. It becomes a question of which effect is larger. When I compute T^* from the fully dynamic model, I see that the benefit of forward exclusion dominates the negative effect of cumulative innovation by competitors for the firm, and T^* decreases.

It is worth noting that in neither model is the welfare maximizing T^* equal to the innovation maximizing T^* , where the rate of innovation is taken to be $\mathbf{E}[\mu_{t+1} - \mu_t]$. I demonstrate this for the dynamic model below.

I conclude this section with an examination of the role of entry costs $\bar{x}^e$ in figure 1.19. In the static and dynamic setting. In the static setting, we mechanically expect the optimal T^* to increase in fixed costs of entry: to incentivize innovation when entry is more costly, firms must be rewarded with larger T . Examination of the equilibria in the dynamic case may lead us to doubt this: entry primarily occurs when there is no active patent, and higher entry costs serve as an effective deterrent to entry when there is an active patent (and forward exclusion on innovation). In fact, I do find that in the dynamic setting, the optimal patent duration weakly decreases in entry

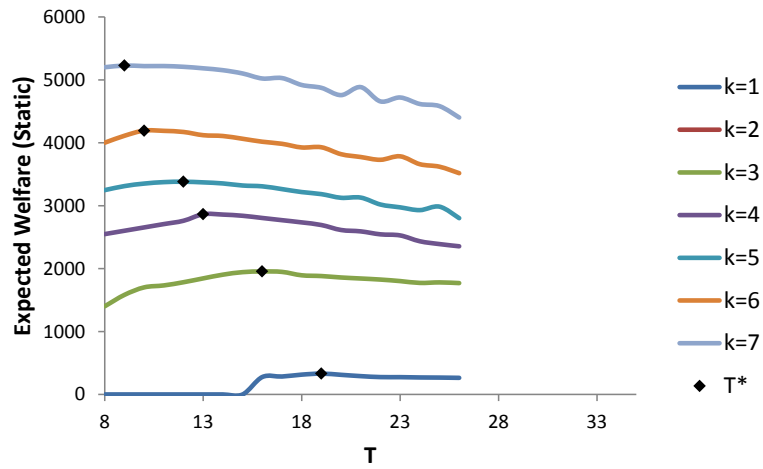


Figure 1.16: Dynamic Welfare as function of T^*

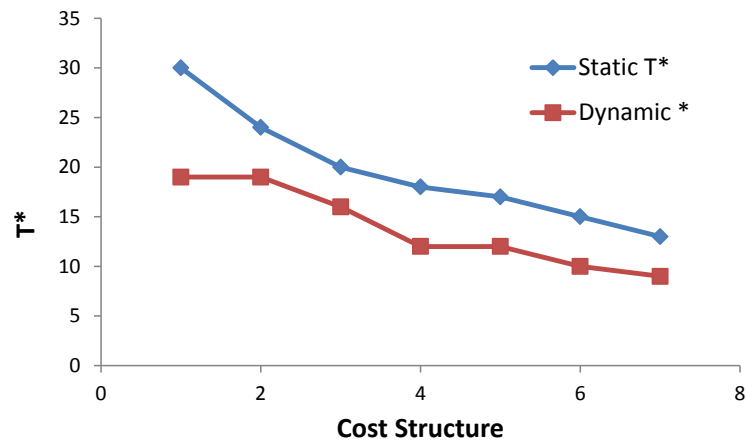


Figure 1.17: Static vs. Dynamic T^*

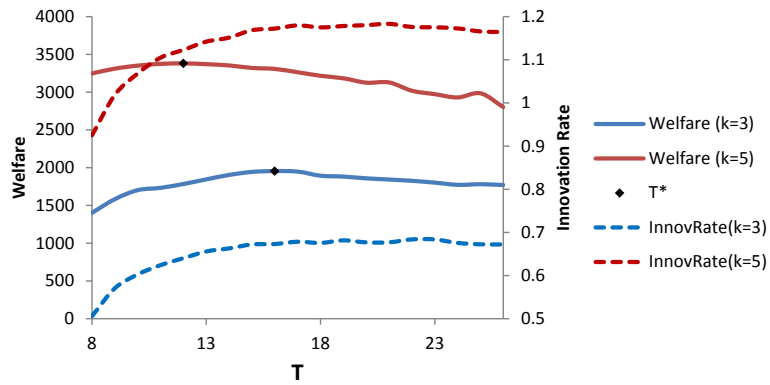


Figure 1.18: Welfare and the Innovation Rate

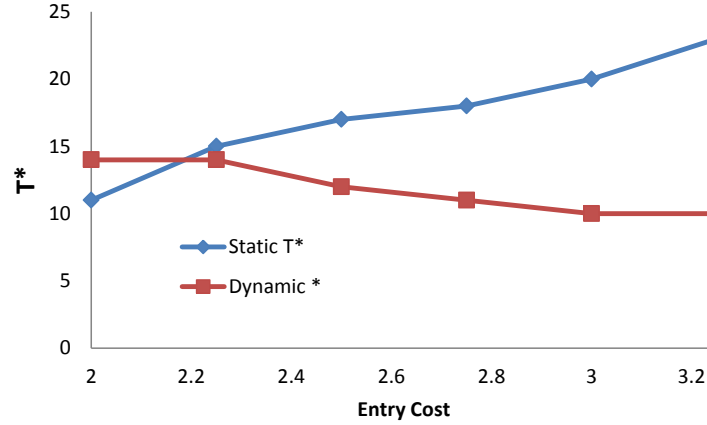


Figure 1.19: Entry Costs: Static vs. Dynamic T^*

costs. The increased value to holding a patent from depressed entry imposed by higher $\bar{x}^e$ tomorrow effectively compensates firms today for the higher cost of entry. I focus on the case $k = 5$, that is $\alpha = .24$ and $\bar{x}^f = .28$, and $\bar{x}^e \in [2, 3.25]$.

I now switch entirely to the dynamic model. The baseline specification unless otherwise noted is as in A.2.

First, I allow market size to vary over five levels, $M \in \{8, 10, 12, 14, 16\}$, and vary fixed costs of R&D between $\bar{x}^f \in \{.25, .4, .55\}$. I expect $\frac{dT^*}{d\bar{x}^f}$ to be positive as in my earlier experiments, but the $\frac{dT^*}{dM}$ is not a priori clear. M shows up in the firms problem (higher M means higher profit levels and higher patent values) but M also shows up on the welfare side (larger M implies larger welfare gains from innovation). In figure 1.20 I see that the first effect dominates. Larger M compensates innovators through larger profits allowing welfare to be improved by lowering T .⁷¹

1.4.2.2 Optimal Patent Duration and Forward Protection

I now consider two policy instruments simultaneously, T , the statutory patent duration and $\tilde{\chi}$ from equation (1.4), which is the automatic “weight” given to the patentee in court proceedings. It can be thought of as a reduced form measure of infringement protection given to past innovators in the

⁷¹Note that because of the linearity of utility and the linear markups it generates, varying M has comparable effects to varying $\mathbf{E}[\lambda_i]$ for both the firms and consumers.

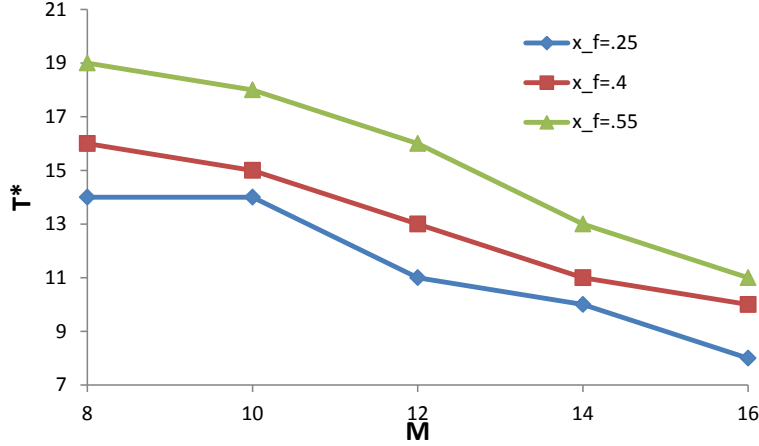


Figure 1.20: Optimal T^* as function of $\bar{x}^f$ and M

court. I solve for T^* and $\tilde{\chi}^*$ jointly. Increased $\tilde{\chi}$ is both pro-competitive (because it increases the value of future patents) and anti-competitive (because it reduces the value of current innovations).

I first examine optimal T^* and $\tilde{\chi}^*$ as a function of the cost settings I examined in 1.4.2.1. There I saw the intuitive relationship, $\frac{dT^*}{dR\&D\ cost} > 0$ for fixed $\tilde{\chi} = 0$. I consider the cost settings

$$k = \{1, 2\}, \quad \alpha_{(k)} = \{.16, .28\}, \quad \bar{x}_{(k)}^f = \{.26, .36\}.$$

In figure 1.21 I have surface plots for welfare as a function of T and $\tilde{\chi}$. I allow $T \in \{5, \dots, 35\}$ and $\tilde{\chi} \in [0, .6]$.⁷²

There are two key takeaways from figure 1.21:

1. *The direction of optimal policy can change depending on which instruments are being optimally calculated.* Recall that for the high- and low-cost environments above, when $\tilde{\chi} = .15$ I calculated optimal T^* of 16 and 10, respectively. Note that when I allow $\tilde{\chi}$ to vary, this flips. The optimal policy in high-cost setting is $(T^* = 13, \tilde{\chi}^* = .26)$. In the low cost setting,

⁷²Even in this homogeneous model, this grid over policies and the three cost environments involves simulating close to 1700 equilibria, or close to 500 million sample “years”. Computing sample statistics for each equilibrium takes about two minutes on average.

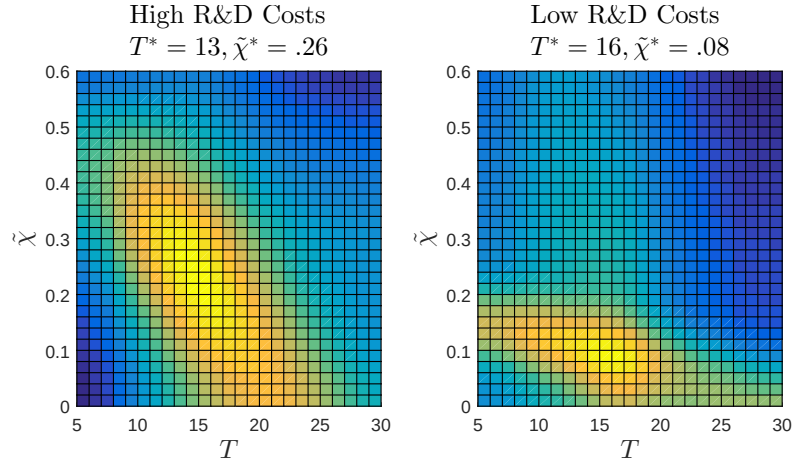
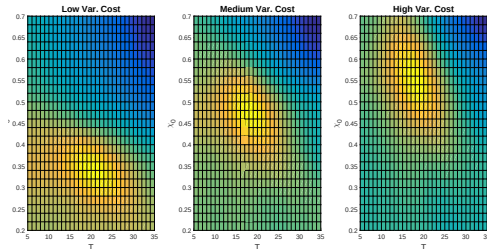


Figure 1.21: Consumer Welfare as function of T and $\tilde{\chi}$.

($T^* = 16, \tilde{\chi}^* = .08$). Note then that the sign on $dT^*/d[\text{R\&D cost}]$ has flipped when I allow other policy instruments.⁷³

2. *Patent Strength is not single-indexed.* That is, T and $\tilde{\chi}$ do not trade off in some one-for-one manner. If this were the case, there would be some optimal level of “patent strength”. So long as $\tilde{\chi}$ and T jointly created this optimal patent strength, there would be some diagonal path through the welfare surfaces through which consumer welfare would be maximized and constant. But I see that $\tilde{\chi}$ and T *do not* trade off in such a manner. (There are definite hints of lower T implying χ to maximize welfare though, as expected.) It follows that forward exclusion strength and statutory duration really are entering the firms problem differently and have different implications for innovation and welfare. They cannot be collapsed into one policy instrument.

⁷³An exercise with a slightly different specification for χ gave similar results. Suppose $\chi_P = \frac{\chi_0 + \ell_P}{\chi_0 + \sum_{j \neq P} \ell_j + \ell_P}$. Computing optimal T^* and χ_0^* in three different cost settings produced a similar result. Fixing $\chi_0 = .4$ yielded $T^* = 15, 20, 23$ for low-, medium- and high-variable-cost settings. Allowing χ_0 to vary yielded optimal policies for low, medium and high-variable-costs of $(T^*, \chi_0^*) \approx (21, .34), (18, .46), (16, .54)$, respectively.



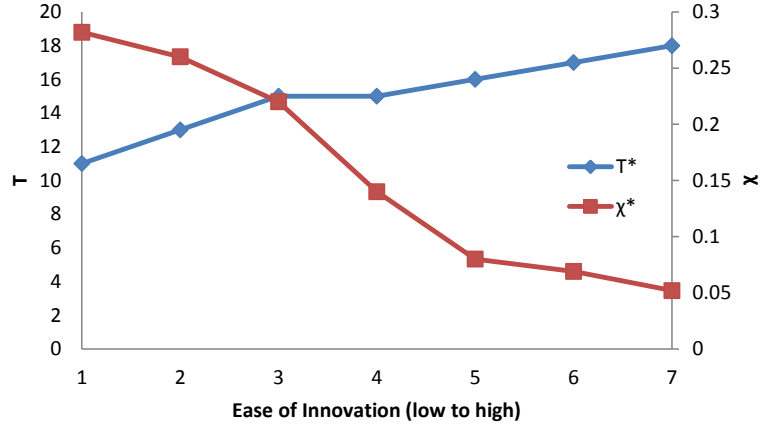


Figure 1.22: Optimal Policy as function of ease of innovation (or R&D costliness)

We see that in lower cost settings, firms are optimally rewarded with less forward protection but more statutory years on the patent. In higher cost settings, firms are rewarded with more forward protection and less statutory duration. My hypothesis is that because costs are convex, it makes more sense for society to reduce competition (and wasted R&D spending) by giving the patentee more forward protection. However, this comes at the cost of lowered statutory duration, forcing the patentee to innovate more frequently to replace expiring patents. In the low cost setting, the welfare cost of wasted R&D effort is less acute. Innovation is cheap and frequent enough to make competition more valuable. This is different to the results in O'Donoghue et al. [1998], where leading breadth should be granted more readily in industries in which is is exogenously easier to innovate. Note though that their leading breadth (essentially, patentability) differs from my litigation-outcome based definition. I report the optimal (T^*, χ^*) for the cost settings in section 1.4.2.1 in figure 1.22.

Consider the low-cost setting from above. In figure 1.23, I examine optimal policy when the exogenous patent-obsolescence rate, ζ , varies. It makes sense that when the probability of a patent being exogenously destroyed increases, it is necessary to reward a patentee with greater forward breadth. The duration instrument has less power, as ζ effectively makes the patentee discount low- τ periods in the future far more. (If the obsolescence rate depicts radically new technologies

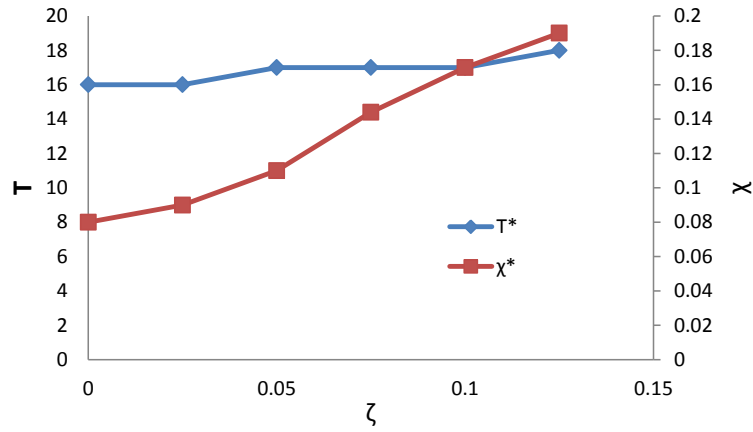


Figure 1.23: Optimal Policy as function of obsolescence rate ζ

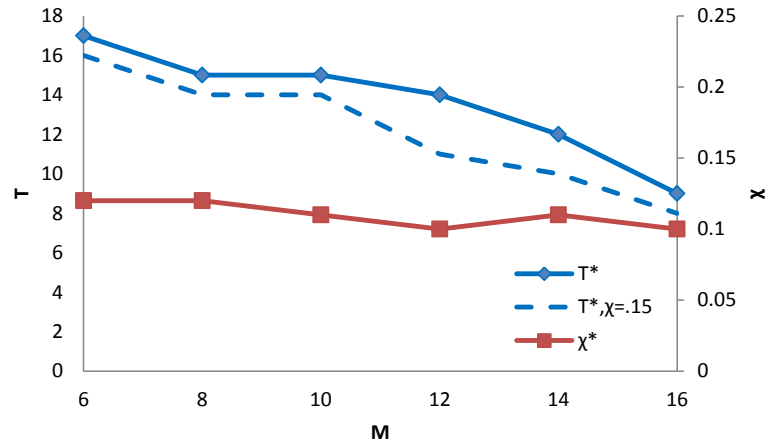


Figure 1.24: Optimal Policy as function of market size M

replacing old patents without any sense of infringement, then this is comparable to the exogenous hit rate in O'Donoghue et al. [1998], perhaps making this the more apt static to compare.)

In figure 1.24, I vary market size M , and allow T and $\tilde{\chi}$ to vary. (I did this earlier in figure 1.20, fixing $\tilde{\chi} = .15$.) The relationship between T^* and M continues to hold, while M has very little effect on $\tilde{\chi}$. (I revise $\tilde{\chi}$ from .15 to $\approx .11$, and T^* increases by between one and two.)

1.5 Conclusion

A flexible but computationally tractable model for examining several different patent policy instruments simultaneously in a fully dynamic oligopoly model of innovation was presented. The paper provides a framework for comparing how patent policy might optimally vary across different industries. In particular, I have taken a first step towards seriously integrating patent litigation and its interaction with patent duration into a dynamic innovation model. My paper is consistent with the reduced form empirical evidence of an inverted **U** shape of innovation and welfare as a function of “patent strength” (Lerner [2005, 2009], Hall and Ziedonis [2001], Galasso and Schankerman [2014], Duguet and Lelarge [2004], Budish et al. [2013]), but makes the theoretical instruments that jointly determine “patent strength” precise. The dynamic model (and the pro- and anti-competitive tradeoffs of patent strength that dynamics entail) is key for generating this, as static models can often give corner solutions for patent policy values.

This paper provides evidence that in industries where market profits are large relative to R&D costs, patent strength should generally decrease. The welfare benefit of reduced rationing and rents outweighs the benefit of increasing innovation with stronger patents. As expected, patent strength in some sense should increase in industries where R&D costs are larger. However, high cost industries are more optimally rewarded via shorter patent duration but stronger forward protection. Industries characterized by quicker technology obsolescence (where technologies change so radically as to leave old patents intact but valueless), firms are optimally rewarded with short duration by strong breadth.

More broadly, the contribution of the paper is that it shows that the policy maker must be careful when they discuss “patent strength”. It is obvious that patent strength is a complicated function of several different policy instruments (duration, infringement protection, etc). What this paper shows is that these instruments do not trade off one-for-one, the function that aggregates them is complicated. Increasing T versus increasing forward exclusion has different implications for innovation and welfare. The direction of optimal policy can change depending on the set of policy instruments that one allows to vary when computing optimal policy.

In the future, I hope to use this paper as a framework for true empirical investigation of this question. A priori it not clear how one could actually observe or infer the deeply unobservable innovations in this model. Hopefully this paper can offer theoretical predictions and guidance on formulating an empirical model. I have constructed a detailed theoretical model with the hope that it will serve as a complement (and not a substitute) to empirical work in this area.

Chapter 2

A Grid Approximation Algorithm for Solving Imperfect Monitoring Games

2.1 Introduction

This paper presents a new algorithm for solving for the perfect public equilibrium payoff set in repeated games with imperfect monitoring and public randomization. These are games in which players repeat a stage game indefinitely, but cannot directly observe the actions of their opponents. Instead, the players must detect deviations from equilibrium play using a publicly visible random signal whose probability distribution depends on the actions taken by the players. The algorithm has some theoretical and practical advantages over the current state of the art algorithm.

The classic example of an imperfect monitoring repeated game is the Green-Porter oligopoly model, in which firms privately choose unobservable quantities Edward J. Green [1984], Porter [1983]. Players must detect deviations from collusive quantities by observing a market price which is a function of both quantities produced by the firms, and a random demand shock. (This model was notably used by Porter [1983] to study a 19th century American railroad cartel.) Repeated games with imperfect monitoring remain of interest to the field and have been used by Athey and

Bagwell [2001] to study collusion in a Bertrand setting, Levin [2003] to solve principal-agent contracting problems, and Radner [1986] to study partnership games.

Despite the field’s interest in games with imperfectly observable actions in topics ranging from firm collusion, monetary policy, insurance schemes and international trade, in general these games remain analytically and computationally intractable for study. Abreu et al. [1990] provide us with a groundbreaking conceptual algorithm (APS) for solving imperfect monitoring games, but implementing the algorithm accurately in a computationally feasible way remains difficult. Abreu et al. [1990] describe a set-valued operator, B^{APS} , whose largest fixed point constitutes the perfect public equilibrium payoff set. In general it is impossible to explicitly construct B^{APS} . The state of the art algorithm for approximating B^{APS} is the Judd-Yeltekin-Conklin (hereafter, JYC) algorithm Judd et al. [2003]¹. The algorithm relies on the fact that convex sets are defined by a set of supporting hyperplanes, and using a subset of these hyperplanes one can construct an outer approximation to B^{APS} by solving a large number of linear programs. (The JYC algorithm is described in detail in section 2.2.2.) The algorithm has several issues. Abreu and Sannikov [2014] show that if the supporting hyperplanes are not chosen correctly, the shape of the resulting set will be qualitatively incorrect. There is no formal guidance for choosing these hyperplanes, nor any relation between the number of hyperplanes chosen and the accuracy of the approximation to the true B^{APS} . There is no convergence guarantee, and if not enough supporting hyperplanes are chosen, the algorithm can “stall” without converging to a set close to the true V^* .

¹Despite publication only in 2003, this paper was known since the early 1990s. The algorithm as it appears in Judd et al. [2003] was designed for perfect monitoring repeated game, but the methodology easily extends to imperfect monitoring. JYC is extended to perfect monitoring stochastic games in Yeltekin et al. [2016]. Cronshaw [1997] extends the JYC methodology to continuous time games. Sannikov [2007] describes methods for imperfect monitoring games in continuous time. Abreu and Sannikov [2014] provides the state of the art algorithm for solving repeated games with perfect monitoring, along with several theoretical insights. Abreu et al. [2016] develops a novel “Twist” algorithm for stochastic games with perfect monitoring. Goldlücke and Kranz [2012] provide methods for solving imperfect monitoring games that allow for monetary transfers. Ericson and Pakes [1995b], Ariel Pakes [1994] provide methods for computing Markov-perfect equilibria in games. Fershtman and Pakes [2012b] provide an algorithm for computing equilibria in games with hidden states and actions, but define a new type of “experience based equilibrium” that differs radically from subgame-perfect and perfect-public equilibrium. Berg and Kitti [2014] provide an algorithm for computing the set of subgame perfect equilibrium payoffs in repeated games *without* public randomization, generating elegant fractals.

In contrast, I construct an algorithm in section 2.2.2.2 that takes a different approach. Rather than approximating B^{APS} with supporting hyperplanes, I construct a grid of boxes over the payoff space. Boxes are then marked as either “feasible” or “infeasible”. A “feasible” box means that a linear program was solved that shows that at the current stage of the algorithm, there is an action profile which is enforceable and which generates a payoff in that box. The convex hull of all feasible boxes is an outer approximation to the APS operator B^{APS} . The accuracy of the approximation is given by the grid resolution: a box of side length κ implies the approximated set is within $\kappa\sqrt{N}$ of the true APS set, where N is the number of players. In practice, testing shows that in many applications the grid algorithm is faster and more accurate than JYC, as shown in section 2.3.

This work is motivated by practical concerns. Out of necessity, analytical work in dynamic games often describe a single Nash equilibrium of a game that has infinitely many Nash equilibria, for example, the trigger strategies of Edward J. Green [1984]. Similarly, in empirical settings, counterfactual analysis is often done using a single computed equilibria. It is unclear if the computed equilibrium used in estimation is robust to the counterfactual policy change. This is why we are interested in computing the full set of equilibrium payoffs. It allows us to make progress towards more robust counterfactual analysis that doesn’t rely on arbitrarily chosen, restrictive equilibria. Furthermore it allows us to ask welfare questions: for example, how far from the Pareto frontier is the observed equilibrium? What is the best equilibrium for society? Additionally, when it is difficult to make assumptions about the structure of the equilibrium that is being played, computing the entire payoff set could help us perform set valued identification of structural parameters as in Abito [2015].

2.2 Algorithm

2.2.1 Theoretical background

This paper considers *repeated games with imperfect monitoring and public randomization*.² These are games where N players indefinitely repeat a stage game. The players cannot directly observe the actions of their opponents, but only a noisy signal that is informative about the actions their opponents took. For example, a low price can be suggestive of a cartel member is cheating, but also a demand slump.

There are players $i = 1, 2, \dots, N$. Each player i has a finite action set $\mathcal{A}_i = \{a^{i1}, \dots, a^{im_i}\}$. There is a *public signal* y drawn from a finite set $\mathcal{Y}$. Using the information in y , players can infer the likelihoods of actions taken by their opponents. The probability that signal y is observed given the players took actions $\vec{a} = (a^1, \dots, a^N)$ is $\pi(y | \vec{a})$. Player i has a stage utility function $u_i(\vec{a}, y)$. In expectation, given action profile $\vec{a}$, player i has stage payoff $g_i(\vec{a}) = \sum_{y \in \mathcal{Y}} \pi(y | \vec{a}) u_i(\vec{a}, y)$. Players have a common discount factor $\delta \in (0, 1)$ and the time- t payoff for player i is given by

$$U_i = (1 - \delta) \mathbf{E}_t \left[\sum_{t'=t}^{\infty} \delta^{t'-t} g_i(\vec{a}_{t'}) \right]. \quad (2.1)$$

These games include a commonly observable random signal z_t .³ This allows for *public randomization*, where players may condition future actions on past public signals.⁴

A strategy for player i is a function σ^i that maps the information set of player i into actions. A strategy is *public* if it depends only on the history of observed public signals, and not a player's own private history of actions. In this paper we will solve for the set of *pure-strategy public perfect equilibrium (PPE)* payoffs. A set of strategies $\{\sigma^i\}_{i=1}^N$ is a PPE if (1) each σ^i is public and (2) given any public history h^t , $\{\sigma^i|_{h^t}\}_{i=1}^N$ is a Nash equilibrium. (In the case of perfect monitoring, PPE reduces exactly to subgame perfect equilibrium (SPE). Like SPE, PPE is a stationary equilibrium.)

²Repeated games with perfect monitoring, as in Abreu [1988], are essentially a simpler special case of these games.

³Without loss of generality Z_t is iid uniform on $[0, 1]$.

⁴Public randomization is what gives us convexity of payoff sets. Without it, we would have fractals as in Berg and Kitti [2014].

We define a “solution” to an imperfect monitoring game as V^* , the set of all payoffs that can be supported by a PPE with public randomization. An *extremal equilibrium* is an equilibrium that corresponds to a payoff that is an extreme point of V^* . Abreu et al. [1990] provide a conceptual algorithm for computing V^* by turning the sequence problem described in (2.1) into a recursive problem. Given a set payoffs $W \subset \mathbb{R}^N$ and that her opponents are playing action profile $\vec{a}^{-i}$, player i can be incentivized to play action a_i if there exists a continuation payoff vector $w \in W^{\mathcal{Y}}$ which maps signals to future payoffs such that

$$a_i \in \arg \max_{\tilde{a}_i \in \mathcal{A}_i} (1 - \delta) g_i(\tilde{a}^i, \vec{a}^{-i}) + \delta \mathbf{E} [w(y) \mid \tilde{a}^i, \vec{a}^{-i}].$$

Each player then has $|\mathcal{A}_i| - 1$ incentive constraints for any action. We say that w *enforces* action a_i . A payoff set W can support a payoff vector $v \in \mathbb{R}^N$ if there exists an action profile $\vec{a}$ and continuation payoff vector $w \in W^{\mathcal{Y}}$ such that (1) w enforces a for all players and (2) $v = (1 - \delta) g(\vec{a}) + \delta \mathbf{E} [w(y) \mid \vec{a}]$.⁵ Note that w specifies a payoff for every player for every realization of the signal y .

Abreu et al. [1990] define the *APS operator* as

$$B^{APS}(W) = \{v \in \mathbb{R}^N \mid W \text{ supports } v\}. \quad (2.2)$$

Abreu et al. [1990] prove several facts about this operator⁶, but most importantly if $W^0 \supseteq V^*$ then repeated application of the operator on W^0 converges to V^* ($(B^{APS})^\infty(W^0) \searrow V^*$).

Conceptually then, the APS algorithm consists simply of choosing a suitably large set W^0 and repeatedly applying the APS operator in (2.2) until convergence to V^* . Practically, it is impossible to analytically apply the APS operator except in simple cases such as the prisoner’s dilemma⁷.

⁵This is called decomposability.

⁶Specifically, the APS operator is (1) monotonic: $W \subseteq W'$ implies $B^{APS}(W) \subseteq B^{APS}(W')$, (2) convergence: if $W^{(0)} \supseteq V^*$ and $W^{(n+1)} = B^{APS}(W^{(n)})$, then $W^{(n)} \searrow V^*$, (3) convexity: with public randomization, B^{APS} maps convex sets to convex sets, (4) self-generation: V^* is largest self-generating set (the largest set where $W \subseteq B^{APS}(W)$).

⁷Even that example is difficult.

Therefore, we require a computational method for approximating this operator in order to study the set of potential outcomes in this valuable class of games.

A special class of imperfect monitoring games are called *repeated games with perfect monitoring*. In these games, players perfectly observe the actions of their opponents.⁸ Although this paper is focused on providing methods for imperfect monitoring games, we compare the algorithm to the state-of-the-art JYC algorithm on perfect monitoring games first. This is because the Abreu-Sannikov (hereafter, AS) algorithm found in Abreu and Sannikov [2014] provides solutions accurate to machine precision.⁹ These serve as a convenient baseline for comparing the accuracy of the solutions provided by grid algorithm and the JYC algorithm. There is no general method for computing exact solutions to imperfect monitoring games at this time.¹⁰ With perfect monitoring, players may condition future actions *directly* on past actions of their opponents (and the public signal).

2.2.2 The Judd-Yeltekin-Conklin (JYC) Algorithm

Judd et al. [2003] provided the first general purpose algorithm for solving these games. Judd et al. [2003] focuses on perfect-monitoring games, but the authors note that the methods can be generalized to stochastic games and imperfect monitoring games.¹¹ The algorithm relies on the fact that any convex set can be defined by its supporting hyperplanes (see figure 2.1). That is, a convex set W can be described as a collection of vectors $\{d_j\}$ and levels $\{\alpha_j\}$ where

$$W = \{x \in \mathbb{R}^N \mid x \cdot d_j \leq \alpha_j \text{ for all } j\}.$$

⁸Basically, $\mathcal{Y} = \times_i \mathcal{A}_i$.

⁹A Java implementation of this algorithm can be found at <http://www.princeton.edu/~rkatzwer/rgsolve/index.html>, with a Matlab front-end at <http://www.princeton.edu/~rkatzwer/rgsolve/JavaRGSolve.m>. Stochastic games may be solved using software at <http://www.benjaminbrooks.net/sgsolvedoc/html/sourcepage.html>.

¹⁰There are other related types of repeated games. *Stochastic games* involve state variables (that may or not be payoff relevant), whose transition probabilities are a function of actions taken in the game. We will not consider these, although both JYC and the grid algorithm can be extended to this case.

¹¹Conklin [1996] and Judd and Yeltekin generalize Judd et al. [2003] for (deterministic) games with state variables.

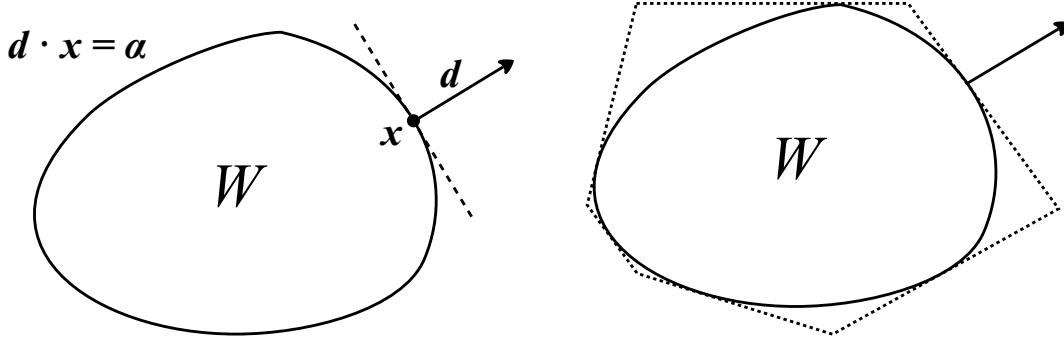


Figure 2.1: Left: A set W and supporting a hyperplane $d \cdot \tilde{x} \leq a$
 Right: A set W approximated by six supporting hyperplanes

The JYC algorithm works as follows: Given an input payoff set W and a search direction $d \in \mathbb{R}^N$, define

$$(w_d, a_d) := \arg \max_{w \in W^{\mathcal{W}}, \vec{a} \in \mathcal{A}} \left\{ d_j \cdot v \mid \begin{array}{l} w \text{ enforces } \vec{a} \\ v = (1 - \delta)g(a) + \delta \mathbf{E}[w \mid \vec{a}] \end{array} \right\}, \quad (2.3)$$

$$v_d := (1 - \delta)g(a_d) + \delta \mathbf{E}[w_d \mid a_d] \quad (2.4)$$

That is, v_d is the payoff that can be supported by W that maximizes the value $\alpha_d := d_j \cdot v$ in the search direction d_j . Given a set of search directions $\mathcal{D} = \{d_1, \dots, d_D\}$, solving (2.3) for each direction d_j generates a set of payoffs $\{v_1, \dots, v_D\}$ and levels $\{\alpha_1, \dots, a_D\}$. JYC then define¹²

$$\begin{aligned} B^{JYC,I}(W) &= \text{co} \{v_1, \dots, v_D\} \\ B^{JYC,O}(W) &= \bigcap_{j=1}^D \{x \in \mathbb{R}^N \mid d_j \cdot x \leq \alpha_j\}. \end{aligned}$$

These are *inner* and *outer* approximations to the true APS operator,

$$B^{JYC,O}(W) \supseteq B^{APS}(W) \supseteq B^{JYC,I}(W).$$

¹²co denotes the convex hull.

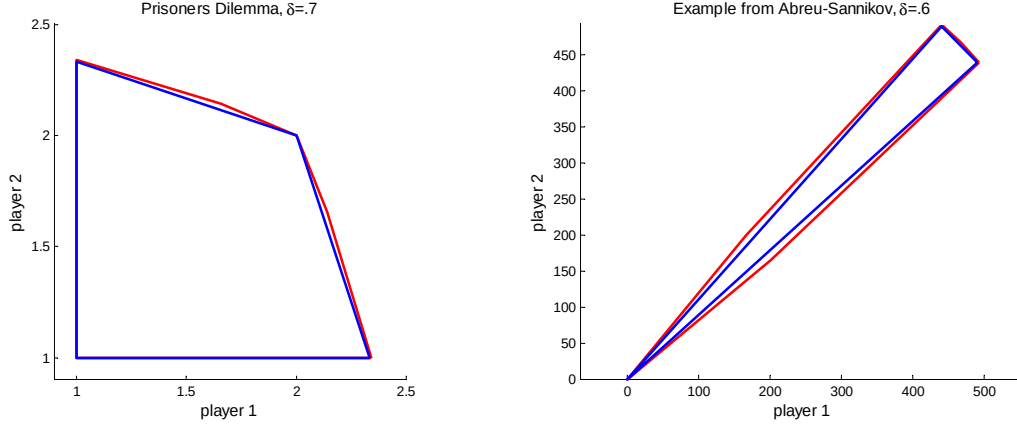


Figure 2.2: Comparison of the JYC set (red) to the true set as computed by AS (blue).

Left: A prisoners dilemma with 32 search directions

Right: An example from Abreu and Sannikov [2014] with 64 Search directions.

The JYC algorithm then repeatedly applies the $B^{JYC,O}$ and $B^{JYC,I}$ operators until convergence. That is, an initial feasible payoff set W^0 is chosen, and then $W^{n+1} := B^{JYC,O}(W^n)$ until some convergence criterion is satisfied.

2.2.2.1 Problems with the JYC Algorithm

Abreu and Sannikov [2014] and others have observed that there are issues with this approach. A general convex set W is potentially defined by an *infinite* number of supporting hyperplanes.¹³ The choice of search directions d_j is arbitrary and *a priori* there is no guidance on which directions to choose. This has implications for the shape of the set that is computed as is shown in figure 2.2. Unless the search directions are chosen to correspond exactly with the normal vectors to the set V^* , JYC will always produce extraneous extremal equilibria. This may seem trivial on these small examples, but without knowing the true shape of V^* it becomes prohibitively expensive to continue

¹³Abreu and Sannikov [2014] prove that in two player repeated games, V^* has a finite number of extreme points that is on the order of the size of the action space. Abreu et al. [2016] provide a similar result for stochastic games. For imperfect monitoring games or games with more than three players, there is no known bound on the number of extreme points.

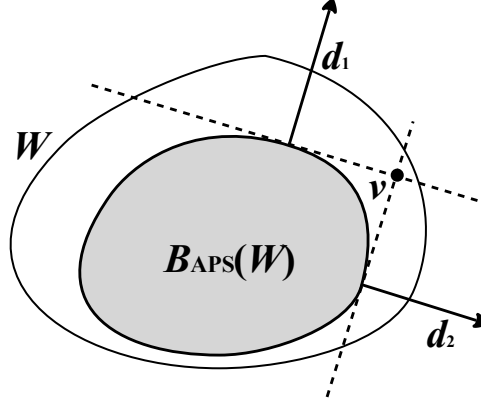


Figure 2.3: Unless JYC uses a search direction in the direction of v , it is possible that v may be retained on every iteration as a payoff supported by W , and the algorithm will “stall” without converging to the smallest bounding polygon of V^* in the search directions specified. There is no way to check for stalling without adding more directions in an ad-hoc manner.

to add search directions.¹⁴ Sets generated by JYC can be quite qualitatively different from the true V^* .

It is impossible to compute a general error band on the distance between $B^{JYC,O}(W)$ and $B^{APS}(W)$. This is because of the presence of extraneous extremal equilibria generated by $B^{JYC,O}(W)$. These equilibria can cause the algorithm to stall before it is close to convergence to the true V^* . There are no results about how the accuracy of the algorithm relates to the number of search directions chosen, or if the algorithm need converge at all as the number of search directions grows but remains finite. Because of these issues, it is difficult to relate the computed $B^{JYC,O}(W)$ and $B^{JYC,I}(W)$ to $B^{APS}(W)$, or the final set to V^* . Without knowing the shape of V^* *a priori* it is difficult to pick the right search directions, or even know how many are needed for a desired accuracy.

2.2.2.2 The Grid Approximation Algorithm

Given the issues with the JYC algorithm, I provide the following algorithm for approximating the APS operator. Rather than trying to maximize supportable payoffs in a set of fixed directions to

¹⁴Adding search directions increases the workload linearly as each search direction must be checked. But less obviously, it also increases the workload exponentially as the linear programs have more constraints. More search directions implies more complicated polyhedrons defining W .

define a set that bounds $B^{APS}(W)$, I instead define a grid over W and solve feasibility problems to plot out the contours of $B^{APS}(W)$. That is, for each box $\mathbf{b}$ in the grid, I check if there is any action profile a and continuation value $w \in W^{\mathcal{Y}}$ such that

$$v = (1 - \delta)g(a) + \delta \mathbf{E}[w(y) \mid a] \in \mathbf{b} \quad (2.5)$$

and every player i is playing a best response given a_{-i} and w . If such a v exists, we say box $\mathbf{b}$ is “feasible”, otherwise “infeasible”. The outer grid approximation to the APS operator is then

$$B^{Grid,O}(W) = \bigcup \{\mathbf{b} \mid \mathbf{b} \text{ is feasible}\}.$$

The inner approximation, $B^{Grid,I}(W)$, is simply the convex hull of the feasible v as in (2.5) that are generated during the course of the iteration. Both sets will be convex polygons. I give the algorithm precisely in 1.

Naively, this might appear to be a horrible approach. Each linear program in (2.6) has 2^N constraints from the hyperrectangle $\mathbf{b}$ ($v \in \mathbf{b}$), $\sum_{i=1}^N |\mathcal{A}_i| - N$ incentive compatibility constraints:

$$(1 - \delta)g_i(a) + \delta \mathbf{E}[w \mid a] \geq (1 - \delta)g_i(\tilde{a}_i, a_{-i}) + \delta \mathbf{E}[w(y) \mid \tilde{a}_i, a_{-i}], \quad \text{for all } i \text{ and all } \tilde{a}_i \neq a_i,$$

and $|\mathcal{Y}|$ constraints corresponding to each linear inequality which defines the set W (that is, the constraint $w \in W^{\mathcal{Y}}$). For a grid with K divisions per dimension, we must solve $K^N |\mathcal{A}|$ many such programs on each iteration. (For comparison, with D many search directions, on each iteration, JYC solves $D |\mathcal{A}|$ many such programs, with each program having 2^N less constraints corresponding to the hyperrectangle $\mathbf{b}$.) As the grid is refined, it is obvious that quite quickly $K^N |\mathcal{A}| > D |\mathcal{A}|$ will occur and the grid algorithm will slow down.

Luckily, we can exploit the properties of the APS operator to significantly cut down on the number of programs we must solve on each iteration.¹⁵ Firstly, since we are only checking if

¹⁵Typically in my experience, the algorithm in practice ends up solving only a fraction of the programs it in theory would need to solve. Less than .005% usually as the grid size grows large.

Define $W^0 = \text{co} \{g(a) \mid a \in \mathcal{A}\}$. Since $\{g(a) \mid a \in \mathcal{A}\}$ is finite, W^0 is a polyhedron defined by some finite number of linear inequalities.

1. Define a grid $\mathbf{B}$ of hyperrectangle over over a hyperrectangle containing W^0 .

(a) Let $W_{temp}^O, W_{temp}^I \leftarrow \emptyset$.

(b) For each hyperrectangle $\mathbf{b}$ in the grid $\mathbf{C}$

i. For each action profile $a \in \mathcal{A}$, try to solve the linear (feasibility) program

$$\max_{w \in \mathbb{R}^{N|\mathcal{Y}|}} 0 \quad (2.6)$$

$$\text{s.t.} \quad \mathbf{b} \ni (1 - \delta)g(a) + \delta \mathbf{E}[w(Y) \mid a] \quad (2.7)$$

$$a_i \text{ a best response to } a_{-i}, \quad \forall i \quad (2.8)$$

$$w(y) \in W, \quad \forall y \in \mathcal{Y} \quad (2.9)$$

If a solution w exists to the program that enforces some action a ,

$W_{temp} \leftarrow W_{temp} \cup \mathbf{b}$ and mark $\mathbf{b}$ feasible and

$$W_{temp}^I \leftarrow W_{temp}^I \cup \{(1 - \delta)g(a) + \delta \mathbf{E}[w(Y) \mid a]\},$$

and break the search over action profiles.

If *no* action profile $a \in \mathcal{A}$ supports a $v \in \mathbf{b}$, mark $\mathbf{b}$ infeasible and continue the search.

(c) Let $B^{Grid,O}(W) \leftarrow W_{temp} \cap W$ and $B^{Grid,I}(W) \leftarrow \text{co } W_{temp}^I \cap W$.

(d) If $W_{temp} = W$, refine the grid.

(e) $W \leftarrow W_{temp}$ and goto step 1b until some stopping criterion is met. (I set a maximum grid size and terminate when W stops changing.)

Algorithm 1: The Grid Approximation Algorithm

there is a feasible payoff $v \in \mathbf{b}$ we need not loop through all actions, we can stop once we find a single feasible action. Moreover, across iterations, if action profile a generated a feasible payoff $v \in \mathbf{b}$ in some previous iteration, action profile a is a good place to start on subsequent iterations when examining the feasibility of $\mathbf{b}$. Since the APS operator is monotonic, once a box $\mathbf{b}$ is rendered infeasible, we can mark it (and its children from subsequent grid refinements) as infeasible forever. We need not check it ever again. If an action a could not support a payoff in box $\mathbf{b}$ with payoff set W , then a cannot support a payoff in any cube $\mathbf{b}' \subseteq \mathbf{b}$ in any future iteration and that action-box pair need not be checked ever again.

We can also exploit the convexity of the APS operator¹⁶, to mark boxes as feasible. Every time we add a point v to W_{temp}^I , we generate $\text{co } W_{temp}^I$ and any boxes $\mathbf{b}$ that overlap with $\text{co } W_{temp}^I$ can be marked feasible and need not be checked during this iteration.¹⁷ Therefore, carefully choosing which boxes $\mathbf{b}$ to examine and generating a large $\text{co } W_{temp}^I$ quickly significantly cuts down on the amount of work that needs be done.¹⁸ Additionally, at the start of the iteration, the vertices of $B^{Grid,I}(W^{n-1})$ are equilibrium payoffs that we generated previously. We can check whether continuation values associated with these payoffs are still contained in W^n . If so, we can add these points to our inner approximation and trim the number of boxes that need to be checked accordingly.

This process is demonstrated in figure 2.4. The dashed blue line represents W entering the iteration. The solid blue line represents $B^{Grid,O}(W)$ at the conclusion of the algorithm.¹⁹ The red x's correspond to boxes $\mathbf{b}$ that were checked and deemed infeasible during the iteration. Green circles represent boxes $\mathbf{b}$ that were marked feasible by solving linear programs for $v \in \mathbf{b}$. Orange circles represent boxes that were marked feasible exploiting the convexity of B^{APS} . Boxes without a marking were not checked during the iterations, demonstrating how in practice algorithm actually needs to do far less work than one might assume from its definition.

¹⁶It maps convex sets to convex sets.

¹⁷For added speed, we use an online convex hull routine during the iteration.

¹⁸We do this using the Bresenham line algorithm, https://en.wikipedia.org/wiki/Bresenham%27s_line_algorithm.

¹⁹The convex hull of the green points roughly corresponds to $B^{Grid,I}(W)$.

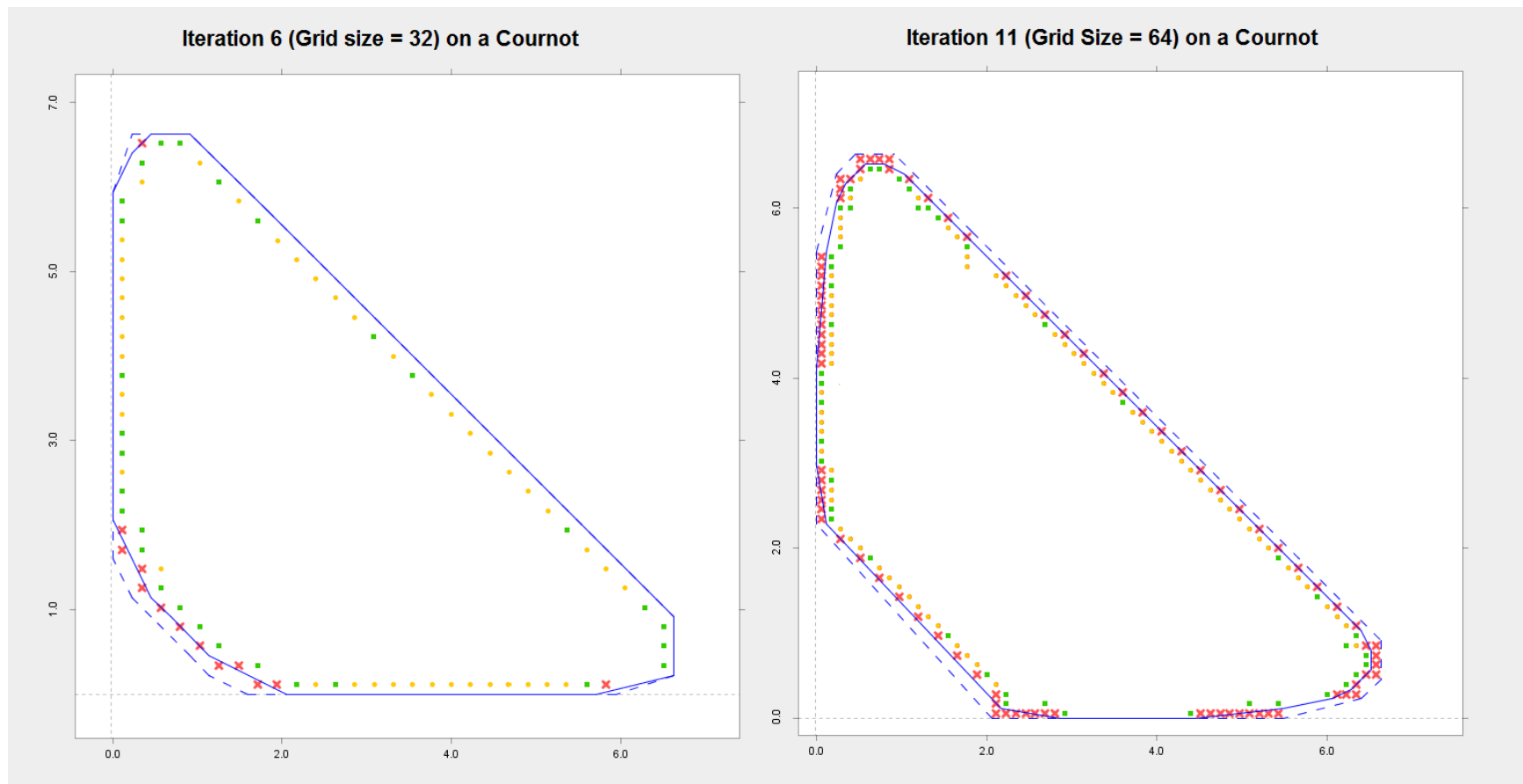


Figure 2.4: Iterations of the grid algorithm on a 2-Player Cournot game with perfect monitoring

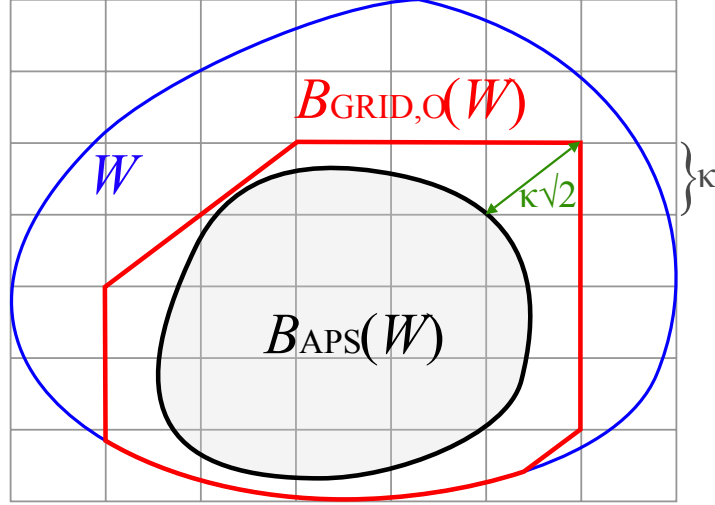


Figure 2.5: The maximum error on the outer grid approximation in a 2-player game.

There are some methods of speeding up the JYC algorithm. For one, when the algorithm is close to convergence, you can fix the action profiles a_d associated with each search direction d (and stop maximizing over all action profiles $a \in \mathcal{A}$), and solve D many linear programs on each iteration, instead of $D|\mathcal{A}|$.²⁰ Additionally, one may “drop” redundant search directions, as the final polyhedron will likely not have D many facets. These tricks speed up the algorithm when one is close to convergence, but are *ad-hoc* and can potentially lead to a less accurate result. Fixing the action profiles potentially breaks $B^{JYC,O} \supseteq B^{APS}$ and dropping search directions potentially increases the size of the final JYC outer approximation and shrinks the JYC inner approximation. None of the methods of speeding up the grid algorithm that were described have any danger of changing the result.

A clear theoretical advantage of the grid algorithm over JYC is that it gives precise error bands on the Hausdorff distance between $B^{APS}(W)$ and $B^{Grid,O}(W)$. Specifically, if each box has side length κ , then the diagonal through the box has length $\sqrt{2}\kappa$. So if the boxes have sides 10^{-3} with 2 players, $B^{Grid,O}(W)$ is within $1.4 \cdot 10^{-3}$ of $B^{APS}(W)$ in the Hausdorff metric. This is demonstrated in figure 2.5.

²⁰Ben Brooks first pointed this out to me.

A more subtle issue with implementing JYC is its generalizability to $N > 2$ players. With 2 players and D directions, one can simply define the search directions as equally spaced vectors around a circle, with adjacent search directions forming a $2\pi/D$ angle. There is no easy way to generalize this to $N > 2$ players. For example, when $N = 3$, except for the platonic solids one cannot equally distribute points on a sphere. Generalizing the grid to $N > 2$ is quite straightforward in theory.²¹

2.3 Results

I implemented the 2-player version of the JYC algorithm and my grid algorithm to compare their performance in practice.²² For demonstration purposes I first present comparisons of JYC and the grid algorithm for perfect monitoring games. The AS algorithm gives us an *exact* solution that can be computed almost instantly. This solution serves as accuracy benchmark.²³ Next, I proceed to imperfect monitoring games, where AS and similar methods are no longer applicable.

In all subsequent tests, given a maximum grid size G the grid algorithm iterated until the grid has been refined to size G and the sets converge so that $W^{n+1} = W^n$. The size of the error, $\kappa\sqrt{2}$ is dictated by G and the size of the payoff space. For perfect monitoring games, the JYC algorithm is iterated until consecutive iterates W^n and W^{n+1} are within $\kappa\sqrt{2}$ of each other (the resolution of the grid, in other words). We don't force an overly stringent convergence criterion on JYC to avoid penalizing it for later iterations which don't contribute much in terms of convergence. Allowing JYC to converge to an error between iterates of say 10^{-4} adds time to JYC without adding much more accuracy relative to the AS computed V^* . I am trying to be as fair to JYC as possible. So in

²¹That said, the memory footprint grows very quickly. This remains the largest issue to solving very fine grids. It is not the amount of work that needs to be done, the issue is storing information about all the boxes we have visited. JYC, by contrast, has almost no memory footprint.

²²The linear programs were solved using the Gurobi optimization suite, Gurobi Optimization [2015]. I wrote all other code in Java, including the convex hull routines, Hausdorff metric calculator, and Bresenham line algorithms. For the AS algorithm, I used the rgsolve code that I wrote for Abreu and Sannikov.

²³JYC and grid algorithm are both obsolete for perfect monitoring games.

these results, for the accuracy of JYC I use a convergence criterion of 10^{-6} , but for JYC's time I record when JYC's error became comparable to the grid resolution in the grid algorithm.

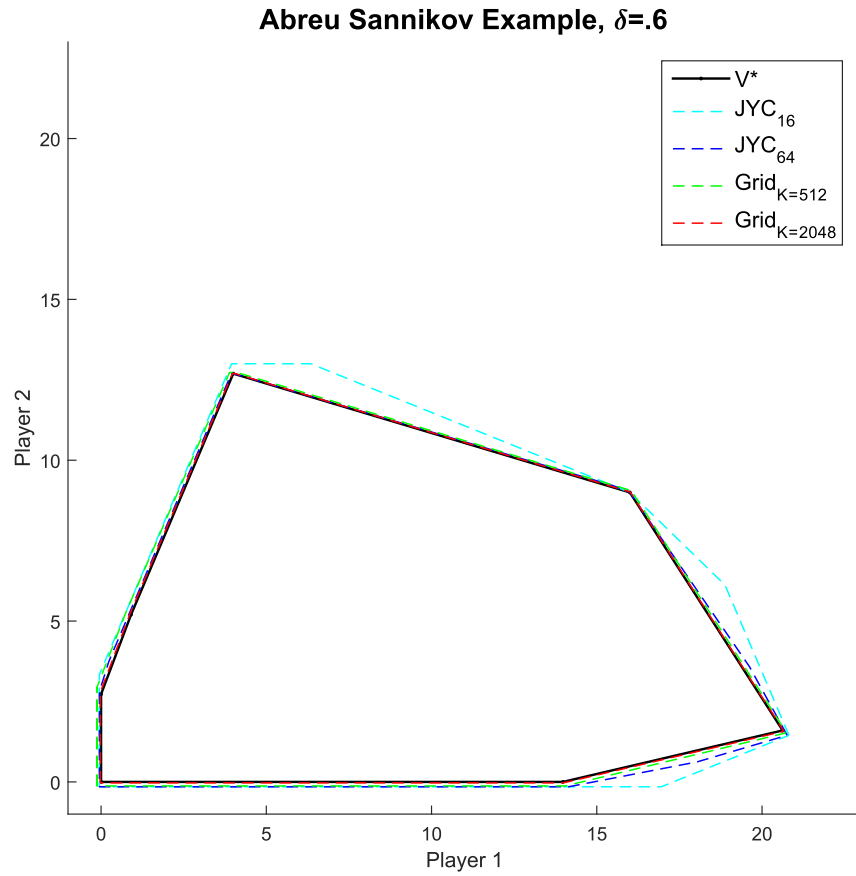
2.3.1 Perfect Monitoring Games

The error in the following tables is the Hausdorff distance between the AS-generated V^* and the approximate V^* produced by the grid and JYC algorithms.

Example 2.1 (Abreu-Sannikov “Simple Example”). Our first example is a Cournot game from Abreu and Sannikov [2014]. The stage game is in table 2.1, and the discount factor is .6. We use a maximum grid size of 1024 and 2048 per player, which corresponds to an error band on the grid algorithm's APS approximation of 0.037 and 0.019, respectively. As one can see in table 2.6, the grid algorithm is far quicker for comparable levels of accuracy as JYC.

	L	M	H
L	16, 9	3, 13	0, 3
M	21, 4	10, 4	-1, 0
H	9, 0	5, -4	-5, -15

Table 2.1: Abreu-Sannikov “Simple Example”



	AS	Grid				AS			
		$G = 256,$	512,	1024,	2048	$D = 16,$	64,	128,	256
Time (s)	.034	.191	.281	.369	.715	.141	.406	.838	2.281
Error	-	.345	.178	.069	.043	.974	.224	.109	.066

Figure 2.6: Abreu-Sannikov “Simple Example”

Example 2.2. The speed advantage of the grid algorithm becomes much more apparent on games with a large action space. Consider an imperfect substitute Bertrand game where player i has demand function

$$q_i = \max \left\{ 10 - 2p_i + \min \left\{ p_{-i}, 10 + \frac{1}{2}p_i \right\} \right\},$$

where marginal costs are $c_1 = .4, c_2 = .2$. Players have a discount rate of .8 and 25 pricing actions.

	AS	Grid				JYC			
		$G = 512,$	1024,	2048,	4096	$D = 16,$	32,	64,	96
Time (s)	.248	7.66	13.4	18.2	32.5	8.91	22.4	102.7	719.2
Error	-	.924	.449	.216	.104	1.74	.926	.851	.328

Table 2.2: 25×25 Bertrand example, results

Notice that when doubling the grid, the time less than doubles (and certainly doesn't square!), but the accuracy more or less doubles as well. There is no such relationship for JYC, where the time grows exponentially for not much reduction in inaccuracy.

Example 2.3 (12×12 Bertrand example). Next, consider an imperfect substitute Bertrand game where player i has demand function

$$q_i = \max \left\{ 18 - 4p_i + \min \left\{ p_{-i}, 10 + \frac{1}{2}p_i \right\} \right\},$$

where marginal costs are $c_1 = .4, c_2 = .2$. Players have a discount rate of .75 and 12 pricing actions. This is a good example of JYC stalling when D is small, as shown in figure 2.7 and table 2.7. With only 20 search directions the resulting JYC set is very inaccurate. Even with 64 search directions, the resulting V^* approximation is far from V^* . The grid algorithm performs quite well, achieving a Hausdorff distance of .114 in 3.1 seconds. Compare this to JYC, which achieves only an accuracy of .341 in 112.4 seconds. For comparable accuracy, the grid algorithm is running about 100 times faster than JYC for this game.

Example 2.4 (A Cournot example). Next consider a Cournot game where the players face the price function $P = \max \{0, 6 - (q_1 + q_2)\}$, marginal cost is .6 and the players have 50 quantity actions over $[0, 6]$. The discount rate is .4.

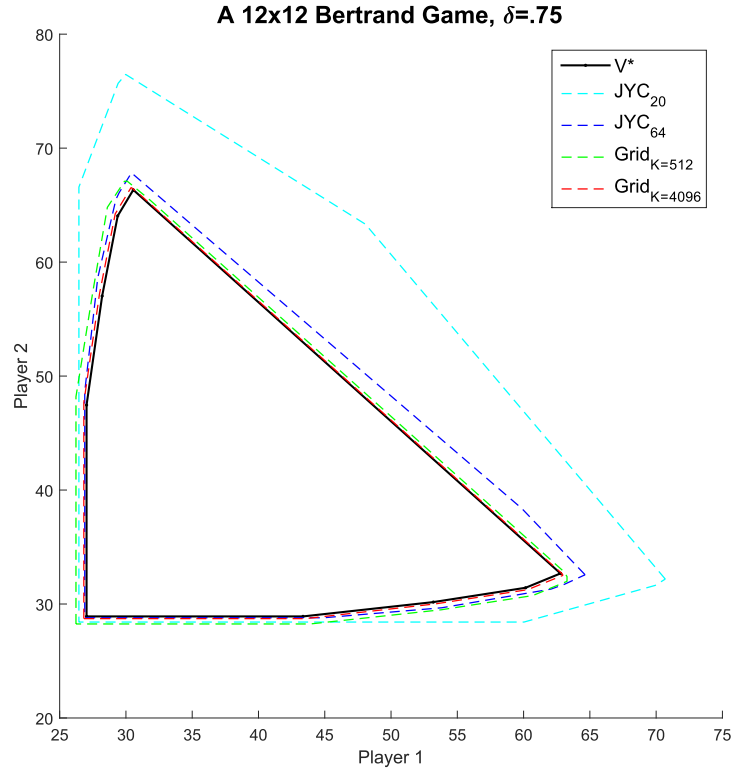
	AS	Grid				JYC		
		$G = 512,$	1024,	2048,	4096	$D = 16,$	32,	64
Time (s)	.145	3.78	7.38	11.5	20.1	3.80	17.67	133.3
Error	-	.026	.017	.010	.006	.168	.092	.013

Table 2.3: 50×50 Cournot example, results

Example 2.5 (A 30×30 arms race). Consider an asymmetric “arms race” where player 1 and 2 can choose an armament level $a_1 \in [0, 12]$ and $a_2 \in [0, 8]$. The payoff functions are

$$\begin{aligned} g_1 &= 3(a_2 - a_1) - \frac{1}{2}a_1^2, \\ g_2 &= 2(a_2 - a_1) - \frac{1}{2}a_2^2. \end{aligned}$$

Each player has 30 actions. We repeat this exercise for $\delta = .5$ and $\delta = .9$. JYC is more competitive at high discount rates in this example.



	AS	Grid				JYC			
		$G = 512,$	1024,	2048,	4096	$D = 20,$	64,	128,	256
Time (s)	.120	.760	1.08	1.88	3.13	.344	7.96	31.7	112.4
Error	-	.978	.449	.265	.114	11.73	1.83	1.67	.341

Figure 2.7: 12×12 Bertrand Example

2.3.2 Imperfect Monitoring Games

When we consider to games with imperfect monitoring, we no longer have the AS-generated set V^* to use as a benchmark for accuracy. However, if $W \supseteq V^*$ then $B^{Grid,O}(W), B^{JYC,O}(W) \supseteq V^*$ since both are outer approximations. So we can compare the algorithms on the size of the sets produced and the amount of time the algorithms used to generate those sets. A smaller set is a more accurate set.

	AS	Grid				JYC		
		$G = 512,$	$1024,$	$2048,$	4096	$D = 16,$	$32,$	64
Time (s)	.131	4.29	6.27	12.1	23.79	1.96	8.49	48.38
Error	-	.242	.138	.074	.033	.851	.247	.087

Table 2.4: 30×30 arms race, results, $\delta = .5$

	AS	Grid				JYC		
		$G = 512,$	$1024,$	$2048,$	4096	$D = 16,$	$32,$	64
Time (s)	.330	20.6	30.7	49.5	88.2	7.65	39.7	233.4
Error	-	1.46	.783	.372	.192	1.34	.384	.122

Table 2.5: 30×30 arms race, results, $\delta = .9$

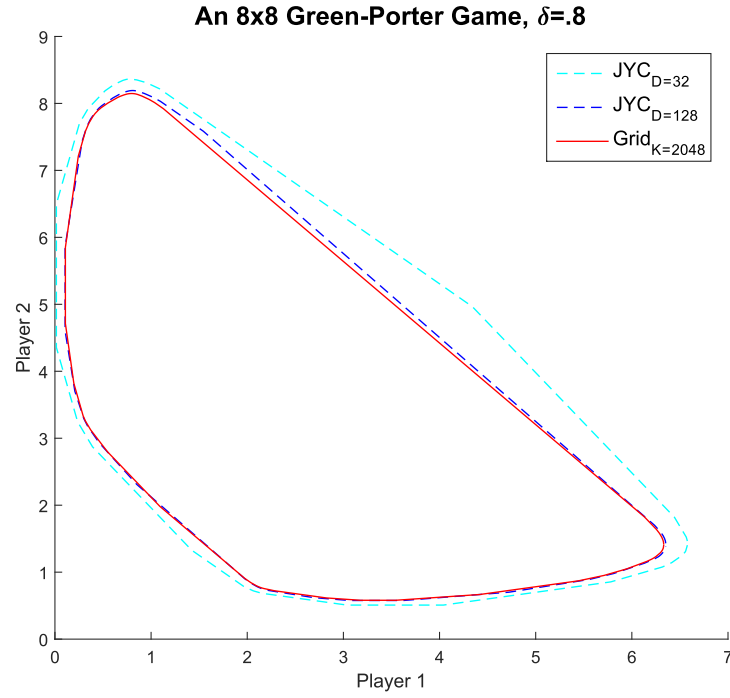
Our convergence criterion for JYC is that successive iterates are within 10^{-6} of each other in Hausdorff distance. For the grid algorithm, we require that $B^{JYC,O}(W^n) = W^n$ on a 2048×2048 grid.²⁴

Example 2.6 (Green-Porter). Consider a game like that in Edward J. Green [1984], Porter [1983]. Players choose quantities and face a joint demand curve

$$P = \max \{6 - (q_1 + q_2) + \varepsilon, 0\},$$

where $\varepsilon \sim \mathcal{N}(0, 1)$. Players do not observe (a_1, a_2, ε) but only the price P , and must detect deviations from this signal alone. Marginal costs are $c_1 = 1, c_2 = \frac{1}{2}$ and the discount rate is .8. Each player has 8 quantity actions on $[0, 6]$.

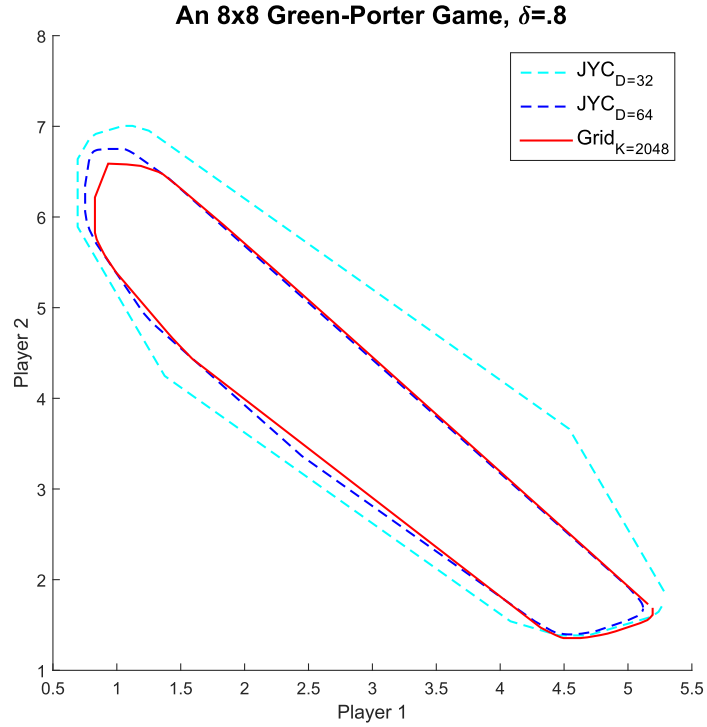
²⁴This corresponds to an error of about $5.5e-3$ in the APS operator approximation given the sizes of the payoff space.



	Grid	JYC	
	$K = 2048$	$D = 32,$	128
Time (s)	139	83	408

Figure 2.8: 8×8 Green-Porter, $\delta = .8$

Example 2.7 (Green-Porter). This example is the same game as above, but with 15 actions per player and a discount rate of .6;



	Grid	JYC	
	$K = 2048$	$D = 32,$	64
Time (s)	112	75	187

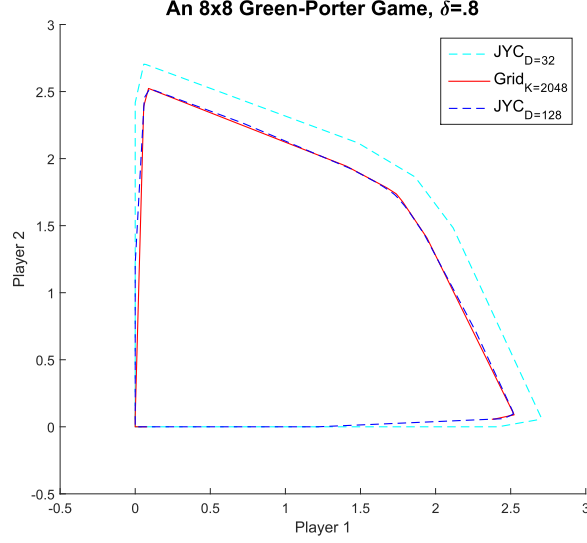
Figure 2.9: 15×15 Green-Porter, $\delta = .6$

Example 2.8. Lastly, I include a classical prisoner's dilemma²⁵ with imperfect monitoring. The game is given in table 2.6.

	Cooperate	Defect		y_1	y_2	y_3
Cooperate	2, 2	-1, 3	(C,C)	.34	.33	.33
Defect	3, -1	0, 0	(C,D)	.1	.1	.8
			(D,C)	.1	.8	.1
			(D,D)	.8	.1	.1

Table 2.6: Prisoner's dilemma: payoffs and public signal probabilities

²⁵This example was given to me by Ben Brooks.



	Grid	JYC	
	$K = 2048$	$D = 32, \quad 128$	
Time (s)	9.4	3.6	14.2

Figure 2.10: Prisoner's Dilemma, $\delta = .6$

Unfortunately, while the grid algorithm performs well against JYC in the imperfect monitoring setting, it is not as dominant against JYC as it was in the perfect monitoring setting. In both algorithms, the number of linear programs to be solved grows roughly on the size of the action space, but the grid algorithm only needs to check a fraction of action profiles for each box while JYC must check them all. This is true in both monitoring settings. The difference though is that with perfect monitoring, the individual linear programs do not grow more complex with more actions since players face only one incentive constraint per action, and continuation values are still in $\mathbb{R}^2$. However, with imperfect monitoring a player i faces $|\mathcal{A}_i| - 1$ incentive constraints for each deviating action she can take. Additionally, continuation values are no longer in $\mathbb{R}^2$ but are in $(\mathbb{R}^2)^{\mathcal{Y}}$ when we switch to imperfect monitoring.

2.4 Conclusion

This paper has presented a new algorithm for solving repeated games with imperfect public monitoring, a class of games with wide interest in the field. Specifically, the algorithm provides inner and outer approximations to the theoretical APS operator, B^{APS} . Given a convex set of payoffs W , the algorithm constructs a grid over the payoff space, and solves a series of linear programs in each grid box to plot out an approximation to $B^{APS}(W)$. Refining the grid gives greater accuracy. The approach differs from the state of the art method in Judd et al. [2003] which relies on constructing supporting hyperplanes to the set $B^{APS}(W)$. Theoretically, the algorithm provides an error on the accuracy of the approximation on the order of the size of a grid box. JYC provides no such error band. In testing, the grid algorithm is significantly faster and more accurate than JYC for perfect monitoring games. In imperfect monitoring games, the grid algorithm is competitive with JYC, if not orders of magnitude better.

Chapter 3

Optimal Patent Duration in a Dynamic Oligopoly

3.1 Introduction

Technological leviathans such as Apple, Motorola, and Samsung have spent years waging a multibillion-dollar battle¹ in American and European courts over their conflicting intellectual property claims. Computer firms amass huge banks of patents to use as a foundation for litigation against one another. Industry commentators were shocked when Apple was awarded a patent by the United States on the “wedge-shaped” design of their laptops, Baldwin [2012]. In the less glamorous but more important field of agrotechnology, Monsanto has come under fire for claiming infringement protection not only for the usage of their genetically engineered soybeans, but for all future generations of soybeans that are grown from the seeds, Stohr [2012].

Because of these recent cases and many others in computers, biotechnology, and pharmaceuticals, the scope and duration of patents has begun to be more heavily scrutinized by the public.

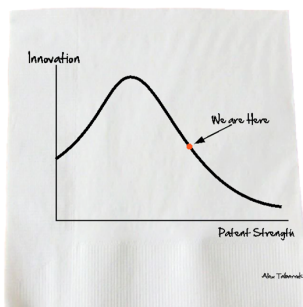
¹For example, a jury awarded Apple a \$1.05 billion settlement against Samsung, based on a patent which the U.S. Patent and Technology Office later declared should not have been granted Lohr [2012]. In November, a British judge ordered that Apple publicly apologize to Samsung (through advertisements) for improperly alleging patent infringement Press [2012]. This lawsuit (to the amusement of the presiding Judge) occurred despite a \$7 billion annual relationship between Apple and Samsung over component manufacture. The press has taken to calling these disputes collectively the “Mobile Device Patent Wars” Barret [2012]

Several economists have weighed in on the issue in the press. For example, Nobel Laureate Eric Maskin writes in Maskin [2012] that in industries characterized by highly sequential progress (in which each innovation builds on the last), “it may be better for society to scrap patents altogether than try to tighten them”. marginalrevolution.com pop-economist Alex Tabarrok presents the prevailing view in Austrian circles: Patent protection is excessively strong and we are on the wrong side of the patent-strength/innovation-rate Laffer curve in Tabarrok [2012].² Michele Boldrin and David K. Levine make an argument for the wholesale elimination of patents in *Against Intellectual Monopoly* (Boldrin and Levine [2010b]), a book intended for non-specialists.

While the weakening or elimination of patent protection might be desirable in some industries,³ it is difficult to make the argument that patents should be eliminated in *all* industries. In industries like pharmaceuticals, the enormous fixed and variable costs of R&D and clinical trials, and the (primarily) brute-force method of molecule discovery make it unlikely that having no patent protection (that is, no monetary reward for innovators) would be socially optimal. With industries such as these in mind, this paper aims to contribute to the optimal patent literature.

What are patents? Before we proceed we should make precise what we mean when we say “patent”. A patent is a form of intellectual property on an innovation or invention. A patent provides “a right...to exclude others from making, using, or selling” the product or process under patent (Rockett [2010]). Under U.S. patent law, many different types of innovations can be

²Tabarrok arrives at the following conclusion using the time-honored Austrian techniques of political conviction and introspection:



³Bessen and Maskin [2009] argue that in industries with sequential innovation and high complementarities to research imitation, patents should be scrapped altogether. They believe the software industry is a good example of a place where intellectual property protection is too strong.

patented; new molecules (pharmaceuticals), an industrial technique or process, a new mechanical invention, an improvement on an old mechanical invention and even certain shapes in industrial design.

Bessen and Maskin [2009] state that the standard economic rationale for intellectual property is to protect innovators from competitor imitation in order to incentivize them to incur the costs of research.⁴ ⁵ Through patent protection, patent holders are granted temporary, artificial monopolist rents leading to deadweight loss and reduced consumer surplus.⁶ The standard story is that patent protection promotes innovation that would otherwise not occur.

Patents are usually characterized by two instruments: statutory patent duration and patent breadth. Patent duration refers to the number of periods that the patent holder is granted monopolist rights to produce/use their innovation.⁷ Patent breadth refers to the size of the “neighborhood” of products around the innovation that are deemed to infringe on the patent and cannot be produced/used.⁸ One should note that *effective* patent duration is what matters; the number of periods the patent holder can actually profit from the patent they hold. This need not be the same as the

⁴This is the patent justification we focus on this paper. A second justification is that the patent system provides an organized disclosure system of new information, in which the new knowledge (if not the right to profit from it) becomes a public good. The ability to act on the knowledge is granted when the patent expires. (Lawyers refer to this as the “teaching” provided by the patent Scotchmer [2006].) Boldrin and Levine [2012] demolish this argument, writing that “from a theoretical point of view... [this justification] fails even the simplest model”. They observe that if patent protection is m periods, then only the innovations that can be kept secret for $n \leq m$ periods will be patented, with the remainder not being patented. So there is no net disclosure gain from a patent system.

⁵Somewhat uniquely, the U.S. Constitution recognizes the right of innovators to profit from their innovation, writing in Article 1, Section 8 that a role of Congress is

To promote the Progress of Science and useful Arts, by securing for limited Times to Authors and Inventors the exclusive Right to their respective Writings and Discoveries.

⁶Legally, “term of patent”. In the United States, patent duration is twenty years from the date of filing, and is legislated in Title 35 Section 154 of the U.S. Code, “Contents and term of patent; provisional rights”. See:

<http://www.uspto.gov/web/offices/pac/mpep/s2701.html#sect2701>

⁷Interestingly, the U.S. Patent Act of 1790 allowed for patent duration to be determined on a case-by-case basis, but not to exceed fourteen years. See: http://www.ipmall.info/hosted_resources/lipa/patents/Patent_Act_of_1790.pdf

⁸One can differentiate between “lagging” and “leading” breadth; lagging breadth refers to product qualities below the state of the art that are too similar to the state of the art and cannot be produced/used. Leading breadth refers to new, superior qualities that are too similar to the state of the art to be patented or used without infringing on the existing patent. A seminal paper on patent breadth is O’Donoghue [1998]. O’Donoghue calculates an optimal “leading breadth”, that is the the quality improvement on the existing state of the art necessary to earn a new patent. He finds that the optimal leading breadth is positive. O’Donoghue concedes the inherent unquantifiability of “patent breadth”. In our model, breakthroughs ($\mu_{t+1} - \mu_t = \Delta$) are effectively granted infinite patent breadth for the duration of the patent.

statutory patent duration that is legislated.⁹ Naturally, patent breadth and statutory patent duration jointly determine effective patent duration.

The United States is characterized by a “one-size-fits-all” patent policy.¹⁰ Software development and bicycle manufacture are industries characterized by different rates of innovation, cost structures, entry barriers, etc., but the same patent policy is applied to both.¹¹ Patent duration is defined through legislation (and is invariant across industries), and patent breadth is subjectively defined on a case-by-case basis by the U.S. Patent and Trademark Office and by the courts.

Scotchmer [2006] identifies two common criticisms from economists regarding the “one-size-fits-all” approach. She writes that the courts and the Patent and Trademark Office “have no mandate to make decisions based on economic reasoning. It is almost inevitable that some classes of innovations are under-rewarded and others are over-rewarded relative to the cost of innovation”. A second criticism is that “the reward structure cannot be modified according to the market structure in which the innovator operates”.

The objective of this paper. The objective of this paper is to give rough qualitative rules for how optimal patent duration varies across industries with different characteristics. To that end, this paper solves and simulates a simple structural model of innovation in a dynamic oligopoly with finite patent duration. Since patent duration is an explicit policy instrument available to the U.S. Congress, whereas patent breadth is harder to define in practice, I focus exclusively on patent duration.¹²

I find that industries with higher fixed and/or variable research costs should have both higher market concentration and longer optimal patent length. Industries characterized by more frequent “technological process” innovations relative to “product quality” innovations should have shorter

⁹For example granting an infinite statutory patent duration on vacuum tubes has an effective patent duration of zero periods as the technology is obsolete.

¹⁰With the notable exception of additional FDA requirements for pharmaceuticals.

¹¹Yes, software design (including pseudo-code algorithm descriptions) are patentable;
http://en.wikipedia.org/wiki/Software_patents_under_United_States_patent_law

¹²In some sense, patent breadth is allowed to vary across industries/innovations via “breakthroughs”. See section 3.2.1. Statutory patent duration in our model is the same for all innovations, but breakthroughs have infinite leading breadths (except for the patent holder) and incremental innovations have zero leading breadth.

patent duration. We find that industries in which individual innovations constitute larger “quality” increases should have shorter patent duration (and lower firm concentration), while industries that have slow quality improvement should have longer patent durations (and higher concentration). (We make these imprecise industry characteristics explicit in section 3.2.1.)

Literature Review. The literature on optimal patent duration has the following central observation; while positive patent duration is generally seen as necessary to induce innovation, excessive patent length or protection could stifle innovation as patent holders will sit on their exclusivity rather than innovate. We do not expect innovation rates to be monotonic in patent strength; rather we expect an inverted U shape. This reasoning implies that some finite optimal patent duration should exist in most models.¹³ Consistent with this reasoning are the empirical results of Lerner [2005] who finds (some evidence) that strengthening patent protection increases the rate of innovation when patent protection is low, and decreases the rate of innovation when patent protection is already high.

Early work on patent duration primarily focused on patent design in a single-innovation setting with one innovator and one innovation. These models assume some per-period demand curve for the innovation and time-zero cost of innovation $c > 0$. Nordhaus [1967] was one of the earliest papers. Nordhaus took patent duration T as the sole instrument. He finds the intuitive result that

¹³We say “most” models, as one can construct reasonable innovation models in which strong patents (or any patents) are undesirable. A notable example is the model of Bessen and Maskin [2009]. They show that in a sequential innovation model with spillovers and positive externalities to the original innovator upon further innovations building on the original innovation, as well as the possibility that imitation goods can increase demand for the original good, patents should be almost non-existent. Obviously, the attractiveness of their results would be incredibly industry-specific. We find the more cut-throat competition in our model to be more appealing in general settings. In our model, upon imitation your profits necessarily go immediately to zero. If we have an industry like pharmaceuticals in mind, this is not a bad assumption at all. Yurukoglu [2012] finds that following Congressional reevaluation of Medicare reimbursements for generic drugs in 2005, generic drug production is a fairly unprofitable industry (leading to shortages). Even economists who are considered most hostile to intellectual property such as Michele Boldrin and David Levine, recognize that the large fixed costs of pharmaceuticals justify some sort of government intervention to protect/reward innovators (Boldrin and Levine [2012]). They note that the estimated time-zero cost of developing the average drug is on the order of \$1 billion, while imitation (given the chemical formula) is completely costless. (PhRMA (the trade organization of pharmaceutical firms), estimates that fewer than 30% of drugs brought to the market recoup their R&D costs Scotchmer [2006].)

Similar to the ideas of Maskin and Bessen, Boldrin and Levine [2010a] consider an endogenous innovation model in a quality-ladder framework, and replicate some existing theoretical results, but use diminishing returns over time to existing technology as the driver innovation rather than jostling for monopolist rights. Crucially, their model allows firms to earn profits even in the presence of competition or superior qualities, which our model explicitly rules out.

if granting monopolist rights yields per-period profits $\pi > 0$, then T should be set so that the discounted present value of a profit stream π for T periods equals c . Intuitively, the value of the rents allowed should just cover the cost of innovation. One sees that because of the deadweight loss cause by the monopolist distortion for those T periods, the rate of innovation dynamically will be less than socially optimal. Tandon [1982] extends Nordhaus' model by allowing both statutory patent duration T and the monopolist rent magnitude π to be policy instruments. In general, he finds that it is difficult to determine optimal patent policies without taking a stand on the functional form of the deadweight loss as a function of π . Gilbert and Shapiro [1990] conduct a similar exercise to Tandon [1982]. For fixed T they let $\pi(T)$ denote the minimum per-period reward to the innovator over patent life T that incentivizes innovation.¹⁴ Then social welfare W is entirely a function of T . They arrive at the decomposition

$$\frac{dW}{dT} = \frac{\partial W}{\partial T} + \frac{\partial W}{\partial \pi} \frac{d\pi}{dT}.$$

The terms $\frac{\partial W}{\partial T}, \frac{\partial W}{\partial \pi}, \frac{d\pi}{dT} < 0$ so one cannot sign $\frac{dW}{dT}$, leaving policy prescriptions ambiguous without more structure.

Klemperer [1990] assumes some reduced form demand system and uses both patent breadth and patent duration as instruments; Klemperer finds no clear-cut policy recommendation but patents should either be “long-and-narrow” or “short-and-broad”. Gallini [1992] considers a similar model but deviates from the assumption of one firm in a vacuum and adds competitors who can engage in costly imitation of the innovator. She finds that optimal patents should have (i) long duration (to discourage imitation) when patent breadth is not an instrument and (ii) broad breadth and patent duration set as short as is feasible while still incentivizing innovation when patent breadth is an instrument.

Taken as a whole, there are no clear cut takeaways from the single-innovation optimal patent literature. Optimal patents are generally “long-and-narrow” or “short-and-broad”, but the optimal patent is highly dependent on functional assumptions.

¹⁴In some sense, higher π policy is comparable to a reduced form measure of broader breadth protection.

In the 1990s, economists began expending greater effort at understanding the effects of patent policy on situations with cumulative innovation in which new innovations build on prior discoveries. The earliest models have two-step innovations with a base innovation and a followup. Papers in this vein include Green and Scotchmer [1995], Scotchmer [1996], Denicolo [2000], Hopenhayn and Mitchell [2010], Chang [1995]. These papers mainly focus on the fact that the first innovator has a positive externality on its competitors that it does not internalize. Therefore, patent policy should reflect the social contribution of the first innovator as well as subsequent innovations when rewarding the first innovator to incentivize a socially optimal rate of innovation. The first innovator can be protected via leading patent breadth or licensing rights for future innovations. The stationarity of our paper ensures that there are no positive externalities from innovation and so the division of rents between sequential innovators is moot.

Using the framework in Grossman and Helpman [1991], economists began to examine patent policy in more complicated infinite-horizon dynamic innovation models. Patent duration in the Quality ladder model of Grossman and Helpman [1991, 1993] is infinite, and patent breadth is zero (every innovation is patentable). The authors are more concerned with characterizing innovation in the macroeconomy and are less concerned with individual industries. The paper provided a useful framework for modeling dynamic innovation and quality improvements in a tractable stationary way. Several papers adopted this framework with an eye towards patent policy in individual sectors rather than macroeconomic growth. The seminal paper O'Donoghue et al. [1998] compare two patent regimens: Infinite duration and narrow breadth (so effective patent life is finite due to new innovations) and small duration and broad breadth (so effective patent duration equals the statutory duration). They find that the second policy has shorter effective policy life, which reduces market distortions but also increases R&D costs. Their recommendation is that leading breadth should be granted when the “hit rate” of new innovations can be characterized exogenously as rapid.

The papers discussed thus far set the table for thinking about patents, but are much simpler than the model discussed in this paper. To our knowledge, only two papers are similar to our analysis, Hunt [2004] and Horowitz and Lai [1996]. Both involve crucial restrictions that make their models

analytically tractable but sacrifice realism. As we discuss below, our paper relaxes both of these restrictions for a fuller model.

Models (such as ours) with multiple innovators targeting the same innovation are called “patent races”. Scotchmer [2006] notes that U.S. while policymakers generally view patent races as beneficial (with an eye towards an accelerated rate of innovation), potential inefficiencies are also present; “[patent races] do not aggregate information in the economy efficiently, and they do not necessarily lead to [socially optimal] investment decisions”. This will be a key tradeoff in our model; more firms can correspond to more rapid innovations but also higher fixed and investment costs.

Hunt [2004] builds a dynamic oligopoly model with sequential innovation, but focuses on “patentability”¹⁵ rather than statutory patent duration. He assumes a *passive incumbent*, that is patent holders do not try to innovate.¹⁶ Hunt derives an optimal innovation step below which innovations cannot be patented. Unlike most papers in this literature, Hunt has an endogenous industry structure which is appealing. The fact that effective patent duration is solely a function of innovation success rates and not any exogenous duration schedule removes time as a state variable and makes the analysis far simpler than models in which the clock on patent duration is winding down. It is this construction which makes the value functions simple and allows him to solve the model analytically; his methods would not work in a model with statutory patent duration.

Horowitz and Lai [1996] analytically derive a welfare-maximizing finite patent duration in a dynamic quality ladder model. They show that the welfare-maximizing patent duration is less than the rate-of-innovation-maximizing patent duration.¹⁷ The key difference between Horowitz

¹⁵O’Donoghue [1998] discusses patentability in depth. Similar in some sense to leading breadth, patentability specifies the minimum improvement over existing qualities that can earn a patent. This can be shorter or longer than leading breadth. If we are in a standard quality ladder, the leading breadth may be of size Δ_{LB} , so that if quality μ is the patented state of the art, then qualities $\tilde{\mu} \in [\mu, \mu + \Delta_{LB}]$ infringe on the μ -patent if they are produced. A separate instrument, patentability Δ_P may state that only qualities $\tilde{\mu} \geq \mu + \Delta_P$ can be patented as sufficiently original. If $\Delta_{LB} < \Delta_P$, qualities $\tilde{\mu} \in (\mu + \Delta_{LB}, \mu + \Delta_P)$ can be produced by competitors to the patent-holder, but not patented.

¹⁶This assumption is an extreme limit case of the endogenous result that we find, that patent holders will innovate less than their challengers. Hunt concedes that the assumption is entirely *ad-hoc*.

¹⁷In our model, if we fix the number of firms we can also replicate the result that the welfare-maximizing patent duration is usually less than the innovation-rate-maximizing patent duration. In our paper, the patent duration that maximizes $\mathbf{E}[\mu_{t+1} - \mu_t \mid T]$ is best. This disappears when one allows for firm entry. The single innovator in their model behaves in qualitatively similar ways to the incumbent patent holder in our model.

and Lai [1996] and this paper is that Horowitz and Lai assume a single innovator who only faces competition in production and *not* R&D, whereas this paper has an endogenous number of potential innovators.

Where this paper fits in the literature. To the best of my knowledge, this was the first paper which examines sequential innovation in a fully dynamic oligopoly setting that includes statutory patent length *and* endogenous market concentration. The complexity of our model (and its large, discrete state space) precludes analytical results. Luckily, my model *can* be analyzed numerically. While I cannot provide explicit theorems, the policy implications of the numerics are quite sharp. This paper analyzes a specific parametrization of utility, demand, innovation, etc. Online, I provide computational tools to the research community that allow for examining the effects of patents on general market and innovation process specifications (see appendix B.6).

I feel that examining the patent duration instrument is a fruitful exercise as most of the modern criticism of patents in practice attacks patent breadths that are too generous and patentability requirements that are too lax (Shapiro [2008]). Patent breadth and patentability are fuzzy concepts that can at best be influenced by legislated guidelines. Statutory patent duration, however, remains an unambiguous instrument in the hands of policy makers. It is easily enforced and interpretable.

A roadmap. In section 3.2, I discuss the model, which is reminiscent of both the standard quality ladder models in trade/macro (Grossman and Helpman [1991]) and the dynamic oligopoly framework common in industrial organization (Ericson and Pakes [1995a], Doraszelski and Pakes [2007]). I outline the innovation process, how patents work, the household's problem, and the firm's dynamic investment problem. We briefly discuss our solution concept, symmetric Markov-perfect equilibrium. In section 3.3, we solve for equilibria. We show how to compute optimal patent duration policies. We discuss how the optimal policy varies in the primitives of the model. In section 3.4, I discuss future extensions to this paper with an eye towards empirics. In appendix B.3 I show that a firm has a unique optimal pure strategy given any competitor policy and state (for the model specification used in section 3.3). In appendix B.4, I show that the model satisfies

the assumptions of Doraszelski and Satterthwaite [2007], and a symmetric Markov-perfect equilibrium exists. Appendix B.6 briefly discusses the `optimalpatents` Java package written to solve this model and how it can be customized and used by the reader.

Remark 3.1. There exists a large literature on patent enforcement, patent litigation, patent disclosure/secretcy, unobservable patent quality and value on the part of both patent issuers and innovators, licensing and a host of other interesting practical topics (see Scotchmer [2006], Rockett [2010] for detailed discussions). I abstract away from these concerns in order to have the simplest possible model in which I can still approach the mechanism of interest. I assume that everything is perfectly observable and interpretable to all firms and the patent office.

3.2 The Model

I construct a model with dynamically optimizing firms. A firm earns profits by holding a patent on an innovation. Firms invest in R&D and are awarded patents on new innovations. New innovations arrive stochastically as a function of R&D expenditures. Firms produce varying quality levels of a single production line, with new qualities superseding old. Only the highest quality can be under patent (by at most one firm).

Framework: Quality Ladders. We construct a model that is similar to the ubiquitous “Quality Ladder” framework in Grossman and Helpman [1991] (henceforth GH), with some key departures. A “quality ladder” is a production line of a single (type of) good that has an infinite number of quality levels, $\mu \in \mathbb{N}$. Higher μ levels are associated with higher utility for consumers. Time-0 quality is initialized at some level $\mu_0 \in \mathbb{N}$, and then through the R&D efforts of firms, the “state-of-the-art” quality evolves from level μ_t to level μ_{t+1} (where $\mu_{t+1} \geq \mu_t$).

The major difference between this model and GH is that this model uses finite patent duration, with duration a policy of a social planner. GH assumes infinite patent length, with patents instantly voided by subsequent innovations. I assume finite patent duration on new innovations, with varying degrees of research exclusivity on new innovations (and hence restrictions on the

ability of competitors to innovate and steal patents). (I discuss this in more detail in section 3.2.1). As GH was focused on macroeconomic growth and aggregate innovation whereas I am narrowly concerned with the effect of patent policy on industry dynamics, this is natural.

GH assumes a continuum of firms, which through a law of large numbers gives no aggregate uncertainty about the state from period to period. In this paper, I assume a finite (and small) number of firms competing oligopolistically. This assumption complicates the model by adding non-trivial state transitions as well as richer firm strategy and competition. A small number of firms is also more realistic when modeling intellectual property dynamics as these industries tend to be highly concentrated.¹⁸

I assume discrete time rather than continuous time. This makes matters more difficult, as continuous time quality ladder models always assume that “ $(dt)^n$ ” terms for $n \geq 2$ are “0” and drop out. This allows them to always assume at most one event occurs at any given time t . In discrete time I cannot make such an assumption and must account for all possible state transitions. I also believe that discrete time is more realistic and would make any future calibration/estimation of the model easier.

Unlike GH, I am only interested in one sector (or one product line) of the economy and therefore consider a partial- rather than general-equilibrium model. Therefore, I only focus on one production line rather than a continuum of goods. For example, if I were discussing innovations on hypertension medications, it is not unreasonable to ignore other goods as such a good has very little substitutability.

This paper also departs from the quality-ladder framework frequently used in Ericson-Pakes style dynamic oligopoly papers following Ericson and Pakes [1995a] (henceforth EP). EP quality ladders differ from GH quality ladders. In EP models, every firm is idiosyncratically moving up and down their own quality ladders. In my model, there is one industry-wide quality ladder, and the entire sector moves up the ladder together (but at most one firm “owns” the highest currently achieved quality on the ladder). I also ignore the rich entry-exit dynamics of EP, but do discuss

¹⁸Oehmke and Wolf [2003] show that traditional measures of firm concentration in industries highly reliant on intellectual property tend to underestimate true concentration (their paper focuses on proprietary agricultural technology).

endogenous market concentration (see section 3.2.3.3). I consider the oligopolies I examine to be in “steady-state” with no significant entry or exit.¹⁹ I do a time-zero entry-exit correction, by adjusting firm concentration to be consistent with zero-economic profits in the long run in the industry.

3.2.1 The Innovation Process

3.2.1.1 The Quality Ladder

At time t there is some “state-of-the-art” quality μ_t . The state of the art tomorrow μ_{t+1} has the potential values; $\mu_t, \mu_t + 1, \mu_t + \Delta$ (where $\Delta \geq 1$ is some fixed number).

- $\mu_{t+1} = \mu_t$ denotes that no innovation occurred in period t .
- $\mu_{t+1} = \mu_t + 1$ denotes that at least one firm was successful in having an incremental innovation, but no firms had a “breakthrough”.
- $\mu_{t+1} = \mu_t + \Delta$ denotes that at least one firm had a successful “breakthrough”.

I assume that the pmf on the value $\mu_{t+1} - \mu_t$ is a function of the firms’ investment decisions $I_1, \dots, I_N$ as well as the current state $\mathbf{s}$. We will make this all precise later.

The idea is that an incremental innovation represents some increase in product quality but not a jump in technology or knowledge. The incremental innovator has made a slightly better product, but the knowledge to do so isn’t particularly interesting (or protect-able) and hence anyone can innovate at any time using the incremental innovation as a basis. The product of the innovator is protected, but the knowledge is not. A breakthrough represents some fundamental jump in existing technology/knowledge, and any improvement would necessarily use this improvement (and hence infringe on its patent).

For example, in the pharmaceutical industry, entirely new chemicals are discovered via an expensive sieving method of throwing everything against the wall and seeing what sticks. When a

¹⁹In future work I hope to relax this and allow for full EP entry-exit dynamics.

useful compound is discovered, this constitutes a breakthrough. However, other drugs are created and patented by slightly changing the molecular structure of existing compounds (which are possibly under patent). This would be an incremental innovation. (See Scherer [2010].)

In some ways, “breakthroughs” capture the effects of patent breadth policies. The point here is that the effective breadth of a patent is not determined by the patent office or legislature, but through the subjective case-by-case rulings of the courts. Giving a firm with a breakthrough research exclusivity rights is similar to patent breadth rights in the literature.

3.2.1.2 Patent Policy

Exogenously, a “policy” is a number $T \in \mathbb{N}$ that represents patent duration. An incremental innovation “I” or a breakthrough “ Δ ” can both be patented.

- An incremental patent expires when
 - T periods have passed with no innovation.
 - an incremental innovation occurs by any firm.
 - a breakthrough innovation occurs by any firm.
- A “breakthrough” patent expires when
 - T periods have passed with no breakthrough innovations.
 - an incremental innovation occurs by the patent holder.
 - a breakthrough innovation occurs by the patent holder.

If a patent is terminated by an innovation, that innovation is awarded the next patent. Any expired patented quality/technology/process etc. becomes public domain and can be used by any firm. Note that I have rewarded “breakthrough” innovations with exclusivity in research and production, while incremental innovators are granted only exclusivity in production. I assume that research exclusivity is only in effect when there is more than one year (including the current period) left on

the current patent; if we are going into a period in which the patent is no longer in effect research is not exclusive.²⁰

The effective patent length of a breakthrough is determined by the patent holder. The effective patent length of an incremental innovation is determined by the combination of the decisions of *all* firms.

If multiple firms achieve a breakthrough, the patent is randomly and uniformly assigned between the successful breakthrough innovators. If multiple firms achieve an incremental innovation and no firm achieves a breakthrough, the patent is randomly and uniformly assigned between the successful incremental innovators.

I assume that firms must “register” their innovations. If the currently patented quality was an incremental innovation, any firm will obviously register any new incremental or breakthrough innovation. A firm holding a breakthrough patent will register a breakthrough patent. A firm holding a breakthrough patent may optimally choose not to register an incremental innovation and will instead find it more profitable to ride out their breakthrough patent. Furthermore, firms *must* register innovations in the period in which they occur, or they are lost by next period. (For example, I do not allow a breakthrough patent holder to successfully innovate a breakthrough with τ years left on the patent clock, invest in no R&D for the next $\tau - 1$ years and then register the new breakthrough from several periods earlier immediately upon expiration of their current patent.)²¹

²⁰I assume this purely for numerical reasons; it gives us the sanity check on the computed policies. The optimal investment policy $I^*(\chi, \tau, \iota)$ should be identical for $\tau \leq 1$ and any (χ, ι) . That is, the if there is one or less years remaining on the patent (including the current period), looking forward everyone should behave the same in a symmetric MPE.

²¹The registration assumption is not entirely satisfactory, but it is necessary to avoid the problem of a far more complicated state space to consider. It would be too difficult to allow firms to build incrementally on existing breakthrough patents, or carry undisclosed innovations forward through time. I would have to discuss a richer patent policy regarding how these increments are treated and how the duration clock is reset. The timing assumption is also a weakness; we know that firms will innovate and sit on those innovations, releasing them several periods later based on the status of their old patents and market conditions. (An example is Intel’s staggered release of new processors.) That said, it has long been known that firms will sit on innovations to avoid cannibalizing current product lines. See Igami [2011] for a discussion of cannibalization fears amongst firms. Pakes [1986] explicitly models the period-by-period decisions about patent renewal.

3.2.1.3 The State

From the perspective of a single firm, the state is a tuple (χ, τ, ι) . $\chi \in \{0, 1\}$ denotes whether or not the firm has the current patent (with $\chi = 0$ being “yes”). $\tau \geq 0$ denotes the number of years remaining on the current patent (possibly zero). ι denotes the size of the last innovation and takes the values 1 and Δ . Note that if $\tau = 0$, the state variables χ and ι are meaningless. Though this is a complex state space (in that the transitions are difficult to calculate), it actually isn’t very large. It is considerably smaller than the state spaces in Ericson-Pakes-type models, which explode in the number of firms. The size of our state space is linear in policy T and is invariant to the number of firms N .²²

Remark 3.2. Note that a breakthrough patent holder will register an incremental innovation iff

$$V(0, T, 1) > V(0, [\tau - 1]^+, \Delta),$$

where V represents the value of the current state.

State Transitions The state (χ, τ, ι) can jump to only a limited number of states (χ', τ', ι') . I list these below.²³ The transitions from an incremental patent are quite simple;

$$(\chi, \tau, 1) \rightarrow (\chi, (\tau - 1)^+, 1), (0, T, 1), (1, T, 1), (0, T, \Delta), (1, T, \Delta).$$

It is more complicated for breakthrough patents.

$$(\chi, \tau, \Delta) \rightarrow \begin{cases} \tau \leq 1; & (\chi, (\tau - 1)^+, \Delta), (0, T, 1), (1, T, 1), (1, T, \Delta), (0, T, \Delta) \\ \tau > 1; & (\chi, (\tau - 1)^+, \Delta), (\chi, T, 1), (\chi, T, \Delta) \end{cases}.$$

From any state we can potentially move to (at most) five successor states.

²²The state space has size $4(T + 1)$.

²³We use the notation $(x)^+ = \begin{cases} x, & x \geq 0 \\ 0, & x < 0 \end{cases}$.

3.2.2 Households

At this stage I discuss the households in a very abstract manner. Later in the paper when we run our numerical exercises we will give a precise specification to the household utility. In appendix B.1, we discuss two common specifications in the literature.

We assume there is a representative household with preferences

$$U \equiv \mathbf{E}_0 \sum_{t=0}^{\infty} \beta^t u \left(\{X_t(\mu)\}_{\mu=1}^{\mu_t}, \{p_t(\mu)\}_{\mu=1}^{\mu_t} \right), \quad \beta \in (0, 1) \quad (3.1)$$

where $X_t(\mu)$ is the time t consumption of quality μ , and $p_t(\mu)$ is the time t price of a unit of quality μ . u is some function which orders consumption bundles (and may also be a function of prices). u may depend on μ . Note that μ_t is a random variable.

We assume that given a price vector $p_t(\cdot)$, (3.1) generates demand functions $X_t^*(\cdot \mid p_t)$ where $X_t^*(\mu \mid p_t)$ is the demand for quality μ given prices $p_t(\cdot)$. We assume that X_t^* is “stationary” in the sense that we do not allow it to depend on μ_t , only on the increment $\iota = \mu_t - \mu_{t-1}$.

3.2.3 Firms

There are N ex-ante identical firms. Each firm i maximizes the expected net present value of its profit streams, i.e.

$$\mathbf{E}_0 \sum_{t=0}^{\infty} \beta^t [\text{profits from market} - \text{investment costs}].$$

In equilibrium, N will adjust so that this will be approximately zero.

3.2.3.1 Oligopoly and Monopoly in the Stage Markets

At time t the only available qualities are μ_t and μ_{t-1} . Under my assumptions, not offering qualities $\mu < \mu_{t-1}$ is without loss of generality. All quality levels $\mu \leq \mu_t$ can be produced at constant marginal cost c at time t .²⁴

²⁴After a “breakthrough”, only qualities $\mu \leq \mu_t - \Delta$ can be produced by non-patent holders.

Assumption. *Firms Bertrand-compete in quality μ_{t-1} and quality μ_t if μ_t is not under an active patent. Firms price at marginal cost and profits in those markets are zero.*²⁵

If μ_t is under an active patent, the patent holder acts as a monopolist. He prices differently according to the demand specification and according to $\iota \in \{1, \Delta\}$. We use $\pi(\chi, \tau, \iota)$ to denote profits to a firm in state (χ, τ, ι) . Obviously, $\pi = 0$ when $\chi = 1$.

3.2.3.2 Investment and Innovation

Let $I_i \in [0, 1]$ denote the research intensity of firm i and let I^{-i} denote the research intensity of the competitors of firm i . Research intensity I_i has cost $C(I_i | I^{-i}, \chi, \tau, \iota)$ where C is strictly increasing, differentiable, and strictly convex in I_i .²⁶

A firm successfully innovates with probability $\phi(I_i | I^{-i}, \chi, \tau, \iota)$ where ϕ is differentiable, non-decreasing in I_i and $\phi(0 | \dots) = 0$. I assume that conditional on successful innovation, a firm has a “breakthrough” with probability $\eta(I_i | I^{-i}, \chi, \tau, \iota)$, where η is differentiable and non-decreasing in I_i . Note then that the firm successfully has innovation step of size δ given by

$$\delta = \begin{cases} 0, & \text{w.p. } 1 - \phi(I_i | \dots) \\ 1, & \text{w.p. } \phi(I_i | \dots) \eta(I_i | \dots) \\ \Delta, & \text{w.p. } \phi(I_i | \dots) [1 - \eta(I_i | \dots)] \end{cases},$$

and $\mathbf{E}[\delta | I_i]$ is an increasing function of I_i . We use $Q_\delta(I_i | I^{-i}, \chi, \tau, \iota)$ to denote the probability function on innovation of size δ as a function of research intensity conditional on the state and competitors, etc.

²⁵If $N = 1$, we assume that there is some competitive fringe of producers who can produce off-patent qualities. Otherwise, in the monopolist case, patents would have zero relevance.

²⁶The reason we use research intensity on $[0, 1]$ instead of investment on $[0, \infty)$ is just so that we have a nice compact action set that is invariant to our parameters. This makes the computation later in the paper significantly easier. One can always just have an investment I with cost I , and write $\tilde{\phi} = \phi \circ C^{-1}$, $\tilde{\eta} = \eta \circ C^{-1}$. Moreover, front-loading all the non-linearity into C allows us to have simpler transition probabilities, which is exploited considerably.

3.2.3.3 Remark on Entry: Zero Economic Profit Condition

I abstract away from a formal model of entry and exit that is common in the Ericson-Pakes/Pakes-McGuire literature. In most papers, a cohort of potential firms considers entry in every period based on market conditions, and incumbent firms decide whether to stay or exit. In this paper I choose not to clutter the state space with a variable number of firms in each period. Rather, I assume that we are in some long-run market steady state implied by the patent length T and the firms are fixed.

Formally, I let $x^e > 0$ be the fixed cost of entry that is paid by all firms at time zero. Since firms are risk-neutral and there is only time-zero entry/exit, an entry cost x^e is equivalent to a per-period fixed cost x^f , where $x^f = (1 - \beta)x^e$.²⁷ ²⁸

Let $\mathcal{V}(N)$ denote the expected time-0 present value of a firm when there are N firms in the market. We define $N(T, x^e)$ as the value such that

$$\mathcal{V}(N(T, x^e)) \geq x^e > \mathcal{V}(N(T, x^e) + 1).$$

Henceforth, I write N to implicitly mean $N(T, x^e)$. For large T , $N(T, x^e)$ will ensure that economic profits in the industry are approximately zero.²⁹

Note that x^e sets a lower bound on feasible T ; when $T = 0$ no firm can possibly recoup his fixed cost and $N = 0$. I let $\mathcal{T}_{\min} = \min_{T \in \mathbb{N}} \mathcal{V}(1) \geq x^e$. That is, $\mathcal{T}_{\min}$ is the lowest patent length that can incentivize *any* firms to enter.

One should note that the definition of $N(T, x^e)$ is *not* a zero profit condition, but rather an approximately zero profit condition. Integral constraints on N prevent a true zero profit condition. However, net of fixed costs, the present value of profits relative to the present value of all expenditures in the economy is small enough that we feel comfortable ignoring profits when I calculate

²⁷Either is natural depending on the story we are telling. We could imagine an auto-manufacturer having high per-period fixed costs due to plant upkeep, labor contracts, etc, whereas a software company might have high entry costs (building a brand, etc) but low fixed costs after that.

²⁸The ability to move between x^e and x^f depends on the time-zero entry/exit; if we had firms dynamically entering and exiting, these would have very different implications.

²⁹For smaller T , the integral constraint that N will give us profits that should be accounted for in a welfare analysis.

social welfare according to (3.1). For N exogenous for any T we need some way to account for the profits.³⁰

3.2.4 The Firm's Problem

I use I^{-i} to denote a generic competitor strategy where $I_j^{-i}(\chi, \tau, \iota)$ is the research intensity decision of firm j in state (χ, τ, ι) .

Holding I^{-i} fixed, let $V(\chi, \tau, \iota)$ be a firm's value function. Then V satisfies the Bellman equation

$$V(\chi, \tau, \iota) = \pi(\chi, \tau, \iota) + \max_{I \in [0,1]} \{ -C(I | I^{-i}, \chi, \tau, \iota) + \beta \mathbf{E}[V(\chi', \tau', \iota') | I, I^{-i}, \chi, \tau, \iota] \}.$$

The expectation operator $\mathbf{E}$ really hides the complexity of the state evolution. In Appendix B.2 I write the transition probabilities out in detail.

3.2.5 Solution Concept: Symmetric Markov-Perfect Equilibrium

Markov perfection is a common equilibrium condition in Industrial Organization. A *Markov perfect equilibrium* is a perfect equilibrium in which players use Markov strategies.³¹ Note that any Markov-perfect equilibrium is also subgame perfect as there is a one-to-one correspondence between subgames and states. Moreover, a Markovian strategy is a best response to Markovian strategies by opponents, so a Markov-perfect equilibrium is still a subgame-perfect equilibrium, even when we relax the Markovian restriction on strategies Doraszelski and Satterthwaite [2007].

³⁰This assumption may seem peculiar given the reputation of highly concentrated industries to have high profits. Pharmaceuticals in particular are often singled out for having inordinately high rates of profit. However, when one properly accounts for long term R&D spending, one sees that the rate of profit is only a few points above the prevailing risk-free rate in the market, with much of this spread attributable to a risk premium Office of Technology Assessment [1993]. Scherer writes that “correctly accounting for R&D as a long-lived investment tends to reduce substantially, if not to eliminate altogether, the inference that pharmaceutical companies are on average achieving supra-normal profit returns” Scherer [2010].

³¹Players play optimally in any subgame. (That is, no non-credible threats.) Players use Markov strategies, that is the player's actions are a function only of the current state. They only condition there actions on payoff-relevant variables. See Tirole [1988], Maskin and Tirole [1988, 2001].

I follow Ericson and Pakes [1995a] in giving a definition of a Symmetric³² Markov-Perfect equilibrium that is tailored to our model.

Definition 3.3. A *symmetric Markov-Perfect Equilibrium* is a tuple

$$\mathcal{M}(T) = (\mathbf{s}_0, V(\cdot, \cdot, \cdot), N(T, x^e), I^*(\cdot, \cdot, \cdot))$$

such that

- An initial state for all firms $s_0^i = (\chi_0, \tau_0, \iota_0)$ (which we choose to be $(1, 0, 1)$).
- A value function $V(\chi, \tau, \iota)$ (which implicitly depends on the number of firms N),
- A number $N = N(T, x^e)$ of firms, such that

$$V(s_0^i) \Big|_{N=N(T, x^e)} \geq x^e > V(s_0^i) \Big|_{N=N(T, x^e)+1}.$$

- An research intensity policy $I^*(\chi, \tau, \iota)$ such that for any $\mathbf{s} = (\chi, \tau, \iota)$, $I^*(\mathbf{s})$ solves

$$V(\mathbf{s}) = \pi(\mathbf{s}) + \max_{I \in [0, 1]} \{ -C(I) + \beta \mathbf{E} [V(\mathbf{s}') \mid I, I_{-i}^*, \mathbf{s}] \}$$

where I_{-i}^* denotes that competitors are playing according to the rule I^* .

Claim 3.4. A symmetric Markov-Perfect Equilibrium exists.

See Appendix B.4. I show this for only the specification of ϕ, η, C that is used in our numerical results.

³²It could be interesting to examine asymmetric equilibria as well, particularly to see if any sort of collusive behavior emerges in investment decisions. Equilibrium selection would be an issue if we were to use asymmetric equilibria in our calculation of T^* . Regardless, it would be interesting to see how radically equilibria could deviate from the symmetric ones we calculate. That would give us some qualitative sense in how far we believe outcomes could differ from the symmetric prediction. If we are content to restrict $N = 2$ and to discretize the action spaces, we could calculate asymmetric (and non-Markovian) equilibria using the stochastic repeated games algorithm developed by Dilip Abreu, Ben Brooks and Yuliy Sannikov.

3.3 Numerical Results

The numerical exercise³³ in the following section has three nested steps:

1. In subsection 3.3.1, I solve for the symmetric Markov-perfect equilibrium $\mathcal{M}(N, T)$ for fixed firm concentration N and policy T , and environment parameters $(x^e, \lambda, \text{etc.})$ as discussed in section 3.2.5. I discuss the optimal behavior of firms in this fixed environment and find that firm investment strategies are fairly intuitive.
2. In subsection 3.3.2, I take the primitive parameters as fixed and show how the endogenous market concentration $N(T)$ varies as a function of T . Taking the policy functions from $\mathcal{M}(N(T), T)$, I simulate the model and compute time-zero welfare statistics. I show how welfare varies as a function of T and demonstrate the optimal policy T^* given primitive parameters. In all examined cases, the optimal T^* is finite. Special attention is paid to the quirks of the solutions which owe to integral constraints on N and the difficulty of entry.
3. In subsection 3.3.3, I take our method of computing T^* from section 3.3.2 and show how T^* and $N(T^*)$ vary as a function of the primitives parameters of the model.

All of our numerical exercises use utility specification #1 in appendix B.1. (This is the perfect-substitute utility specification used in Grossman and Helpman [1991] and Horowitz and Lai [1996].) I ran similar simulations for specification #2 (unit demand), and found qualitatively similar results in firm behavior and optimal patent duration.

³³I have written a Java package (optimalpatents) that computes MPEs and simulates them. The user must supply functional forms u, ϕ, η, C, π and parameters $\beta, \lambda, \Delta, x^e, c$. The package can be downloaded at <http://www.princeton.edu/~rkatzwer/optimalpatents/optimalpatentscode.zip>. The solver uses the Brent univariate optimization routine from the Apache Commons math library <http://commons.apache.org/math/>.

In appendix B.6, one can find instructions on how to customize the primitives of this model (and then solve/simulate) using the optimalpatents package.

3.3.1 Policy Functions for fixed N and T

My computational procedure for finding a symmetric MPE is similar to the Gaussian methods discussed in Pakes and McGuire [1994]. We guess a value function V^0 and intensity policy I^0 . Given V^n, I^n I loop through the state space

$$\{(\chi, \tau, \iota) \in \{0, 1\} \times \{0, \dots, T\} \times \{1, \Delta\}\}$$

and compute optimal policies I^{n+1} given the value function V^n and given that competitors play according to I^n . I compute a new value function V^{n+1} and repeat this process until $\|V^{n+1} - V^n\| < \varepsilon$ for some distance metric $\|\cdot\|$.³⁴

The cardinality of the state space grows linearly in T and is invariant to the number of firms. However, the time per iteration grows exponentially in the number of firms, as computing the transition probabilities (as in appendix B.2) becomes very expensive.

A problem in Ericson-Pakes/Pakes-McGuire type frameworks is a potential multiplicity of symmetric MPEs. My MPE solutions do not seem to depend on the initial guess. Additionally, Doraszelski and Satterthwaite [2007] report certain cases in which different symmetric equilibria can be found depending on the iteration scheme; Gauss-Jacobi (new policies are used immediately) or Gauss-Seidel (new policies are used only after the iteration in which they were computed is complete). My MPEs appear to be invariant to the iteration scheme. The results here are from the Gauss-Jacobi computations.³⁵

Below we give an example of an MPE to illustrate firm behavior and the welfare tradeoffs that will be examined in the simulations. I do not solve for the endogenous $N(T, x^e)$ in this example (solving for $N(T, x^e)$ requires the simulations discussed in section 3.3.2).

³⁴The state space is small enough that I can avoid the stochastic approximations as in Pakes and McGuire [2001]. However, if I were to add a richer model of endogenous period-by-period entry and exit or firm-specific productivity, I would likely need to use the stochastic methods, or rewrite the model in continuous time and use the methods of Doraszelski and Judd [2012].

³⁵When $N = 2$, the numerical results lead to the hunch that the Bellman operator here is in fact a contraction and the solution is unique.

Parameters		Functional Specifications
N	4	$\phi(I \dots) = .5I$
T	10	$\eta(I \dots) = .1$
Δ	4	$C(I) = 500 \frac{I^3}{1-I}$
λ	1.1	$U(\{X_t\}) = \sum_{t=0}^{\infty} \beta^t \log(\sum_{\mu} \lambda^{\mu} X_t(\mu))$
β	.95	
c	.1	
$\bar{Y}$	4	

Table 3.1: Parameter Values

(See figures 3.1, 3.2, 3.3, 3.2) I use the parametrization in 3.1. Monopolist profits are $\bar{Y}(1 - \lambda^{-t}) \approx .36, .99$.³⁶ The policy functions below conform with our intuition. A patent-holder on an incremental innovation slowly ramps up innovation as he approaches the expiration of his patent.³⁷ A patent holder on a breakthrough invests far less while his patent is fresh (and he faces no chance of having the patent stolen from him) but begins to invest more as the patent approaches expiration. The value function of the non-patent-holder after a breakthrough is actually *decreasing* in τ , as τ is the number of years that they are effectively shut out of the market. As expected, the value function of a patent holder decreases as τ winds down. A good sanity check on the calculations is that $V(0, 0, t) = V(1, 0, t)$, and

$$I^*(0, 0, t) = I^*(1, 0, t) = I^*(0, 1, t) = I^*(1, 1, t).$$

(Any firm regardless of the state should make the same intensity choice when $\tau \leq 1$, as they all face the same future.)

We see that the firms in our model are very cognizant of the creative destruction dynamics described in Schumpeter [1994 (1942)]. The value of old innovations (along with old innovators) are constantly destroyed by new innovators and innovations. When a firm has a breakthrough, this process slows down tremendously.

³⁶The convergence metric was $\max_{\chi, t, \tau} \frac{|V^{k+1}(\chi, t, \tau) - V^k(\chi, t, \tau)|}{1 + |V^k(\chi, t, \tau)|}$. The computation took 188 iterations to a precision of $1e(-8)$. This took .628 seconds.

³⁷It is not visible but the research intensity of non-patent-holders decreases; as the monopolist ramps up investment the returns to investment by a non-patent holder decrease.

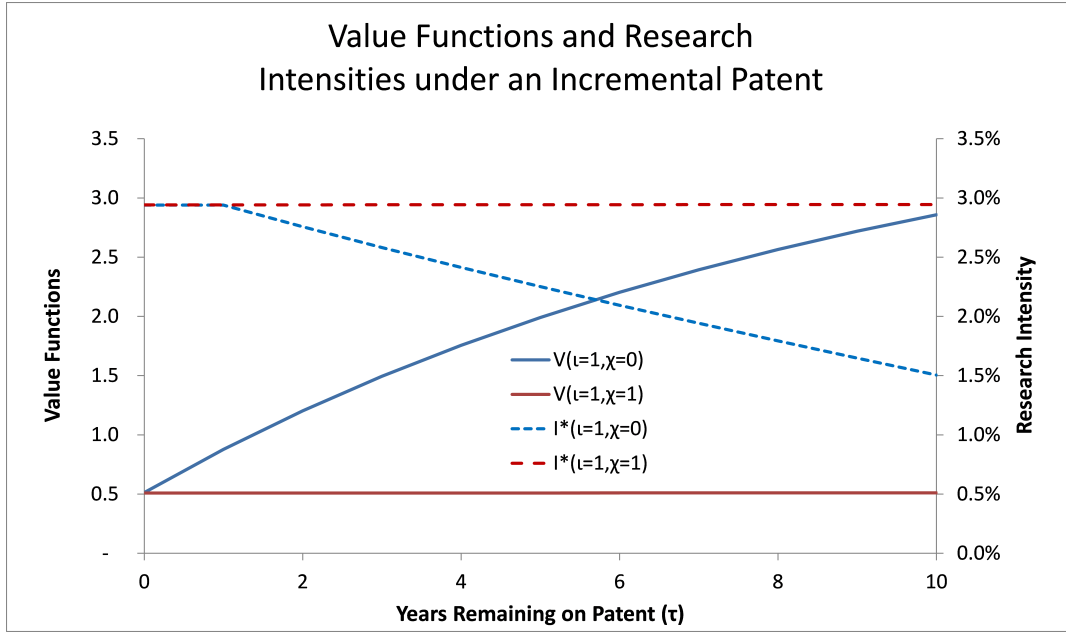


Figure 3.1: MPE, $\iota = 1$

We can compare innovation rates at a given time until expiration τ in the following chart. As expected, innovation rates collapse after a breakthrough as patent holders sit on their research exclusivity. When $\iota = \Delta$, incremental innovations are only registered if $\tau \leq 3$.³⁸ The patent holder is in fact throwing away innovations. The innovation rates are much higher when $\iota = 1$ (dashed lines) than when $\iota = \Delta$ (dotted lines).

In figure 3.4 I show aggregate R&D spending. R&D spending is roughly twice as large after an incremental innovation than after a breakthrough.

³⁸That is, $V(0, T, 1) \geq V(0, \tau, \Delta)$ when $\tau \leq 2$.

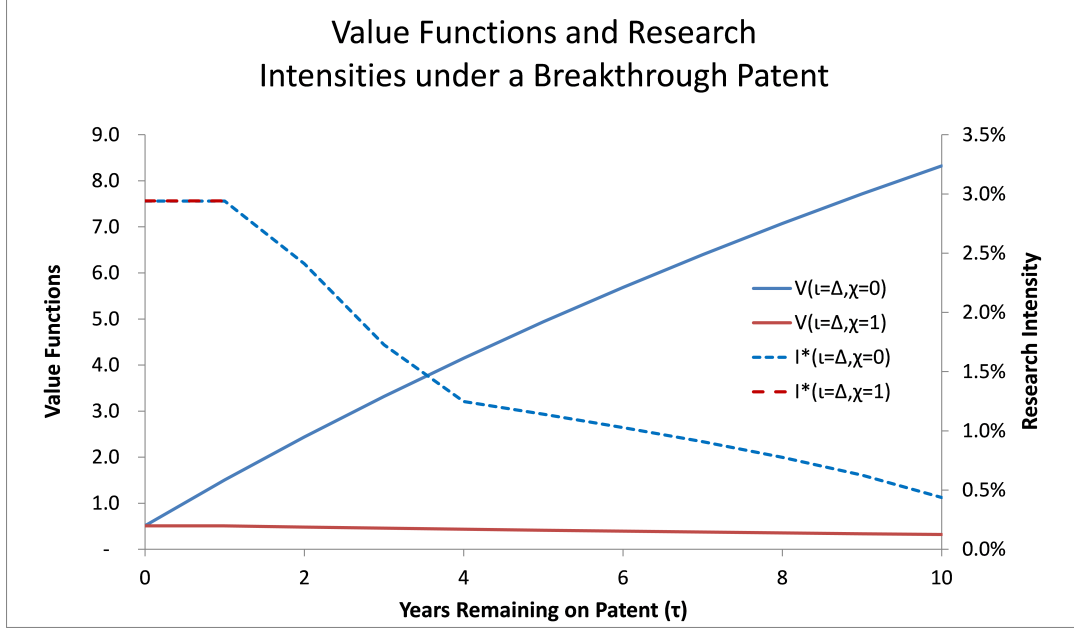


Figure 3.2: MPE, $t = \Delta$

3.3.2 Finding $N(T)$ and T^*

An emphasis of this paper is that firm concentration matters. In appendix B.5 I present an example to show how radically the optimal patent length implications differ when I allow market concentration to endogenously depend on T .

Finding $N(T)$ Given N and T , I can solve for an MPE $\mathcal{M}(N, T)$. Since the initial state for all firms is $\mathbf{s}_0 = (0, 0, 1)$, we note $V_{\mathcal{M}(N, T)}(\mathbf{s}_0)$ (from the MPE) is the expected discounted profit of each of the N firms. I assume that N firms enter so that

$$V_{\mathcal{M}(N, T)}(\mathbf{s}_0) \geq x^e > V_{\mathcal{M}(N+1, T)}(\mathbf{s}_0)$$

and call this value $N(T)$. Note that finding $N(T)$ requires repeated solution of the MPE, but does not require any simulation.³⁹

³⁹Note that I am assuming that $\mathbf{s}_0 = (1, 0, 0)$. It could be interesting to examine $N(T)$ for other starting states $\tilde{\mathbf{s}}_0$. The number of firms who enter at time-zero given state $\tilde{\mathbf{s}}_0$ could give us some clues about how firms would behave with dynamic entry/exit. (I thank Prof. Abreu for this insight.)

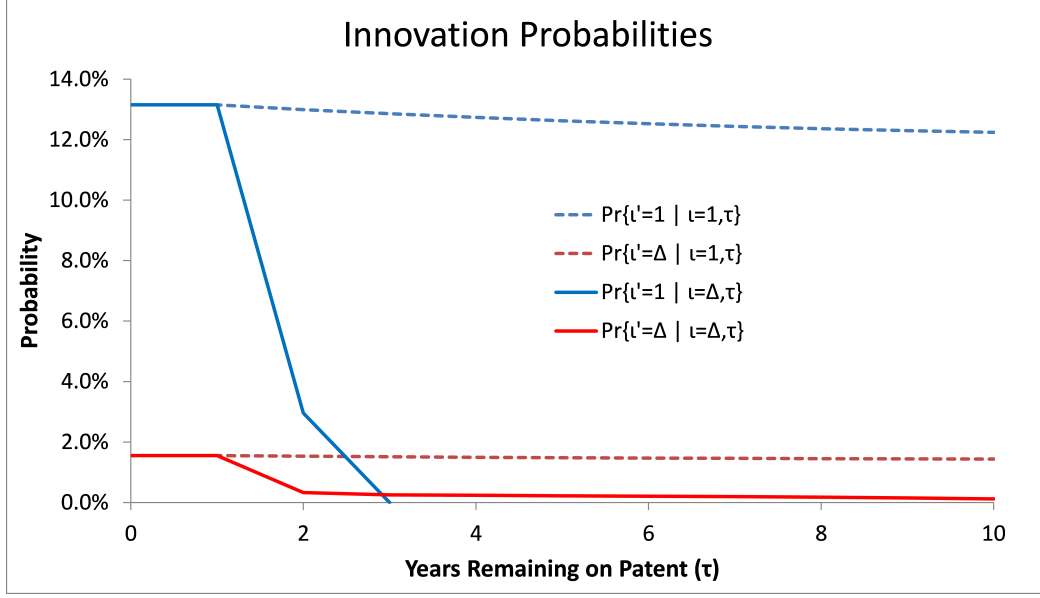


Figure 3.3: Innovation Rates

Finding T^* Solving for $N(T)$ allows us to denote the MPE as $\mathcal{M}(T) = \mathcal{M}(N(T), T)$. So each value of T uniquely determines some equilibrium policy rule from the MPE, say $I_{\mathcal{M}(T)}^*$. Using this policy rule and the transitions discussed in appendix B.2, I can compute time-zero statistics such as the rate of innovation, the discounted present value of profits, R&D costs, fixed cost investments, and consumer welfare via Monte Carlo simulation. I let T^* denote the patent policy that maximized consumer welfare.⁴⁰ I let $\mathcal{T}_{\min}$ denote the lowest patent duration in which firms can recoup their fixed costs of operation and enter.

Unfortunately, the costly simulations are necessary. Even given an optimal policy $I_{\mathcal{M}(T)}^*$, I cannot compute a value function for the consumers, as μ_t would be a state variable rendering the entire exercise moot.⁴¹

Simulations My parametrization for this section is in table 3.2. I choose simple functional forms for ϕ and η since this is the case for which we can establish that a symmetric MPE exists (see appendices B.3, B.4). All of the non-linearity is front-loaded onto a very convex C . I consider two

⁴⁰Since profits are approximately zero, I ignore them and assume that consumer welfare equals social welfare.

⁴¹While the firms' problems are stationary, the period-to-period outlook of the consumer is not. Expected discounted utility explicitly depends on the values of $\{\mu_t\}$. The firms only care about the sequence $\{\mu_t - \mu_{t-1}\}$.

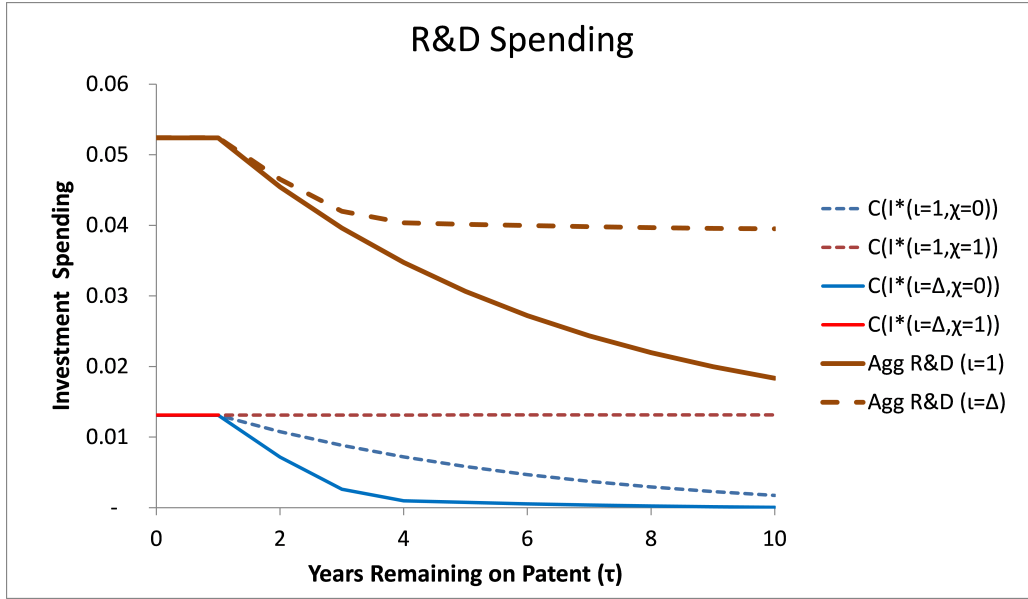


Figure 3.4: R&D Spending

Parameters		Functional Specifications	
Δ	3	$U(\{X_t\})$	$\sum_{t=0}^{\infty} \beta^t \log(\sum_{\mu} \lambda^{\mu} X_t(\mu))$
λ	1.1	$\phi(I \dots)$	$.35I$
β	.95	$\eta(I \dots)$	.2
c	.3	$C(I \dots)$	$500I^4 / (1 - I)$
$\bar{Y}$	5		

Table 3.2: Parameter Values

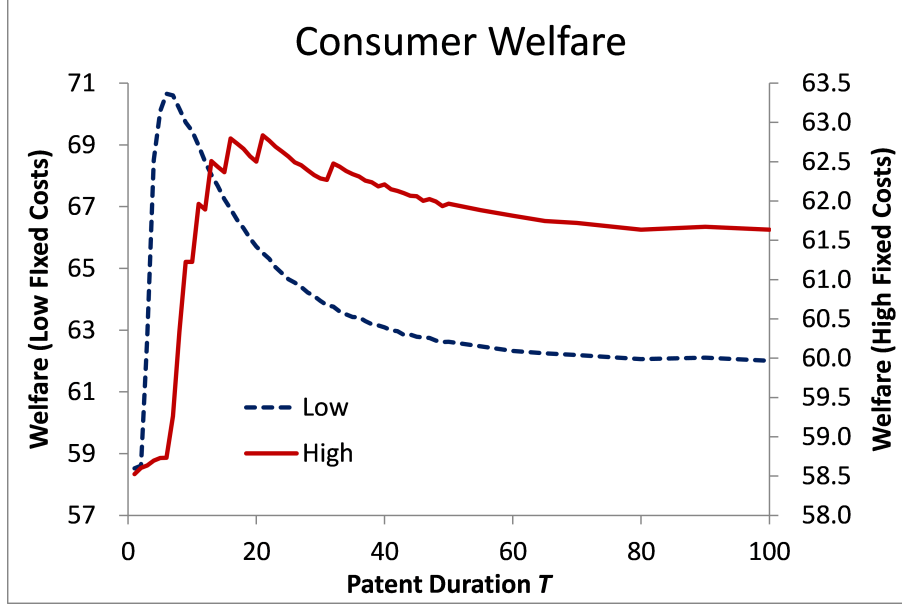


Figure 3.5: Welfare, under high and low fixed costs.

examples corresponding to high- and low-costs of entry and show how the results vary qualitatively. High entry costs are $x^e = 1.35$ and low entry costs are $x^e = .45$. For each T , I simulate 25,000 draws to 500 periods each.

The jaggedness of the consumer welfare curve for high-fixed-costs in figure 3.5 is not a numerical artifact. It corresponds to the observation in the simulations that given any number of firms N , the planner wants to choose T such that T is the smallest duration that can keep N firms in business. Any larger T that does not induce higher entry is giving rents away to the firms. The jaggedness does not appear in the low-fixed-cost consumer welfare curve as the rate of entry as a function of T is so rapid. We generally see here that higher fixed cost industries merit longer patent durations (6 and 21 periods, respectively, for low- and high-fixed-costs). Optimal market concentration is the reverse (20 and 7 firms, respectively, for low- and high-fixed-costs). This is intuitive as incentivizing research through monopolist rewards will obviously be costlier if the fixed costs of operation are higher.⁴²

⁴²Consumer welfare is plotted on separate axes, as there is no meaningful comparison between utility levels between the industries.

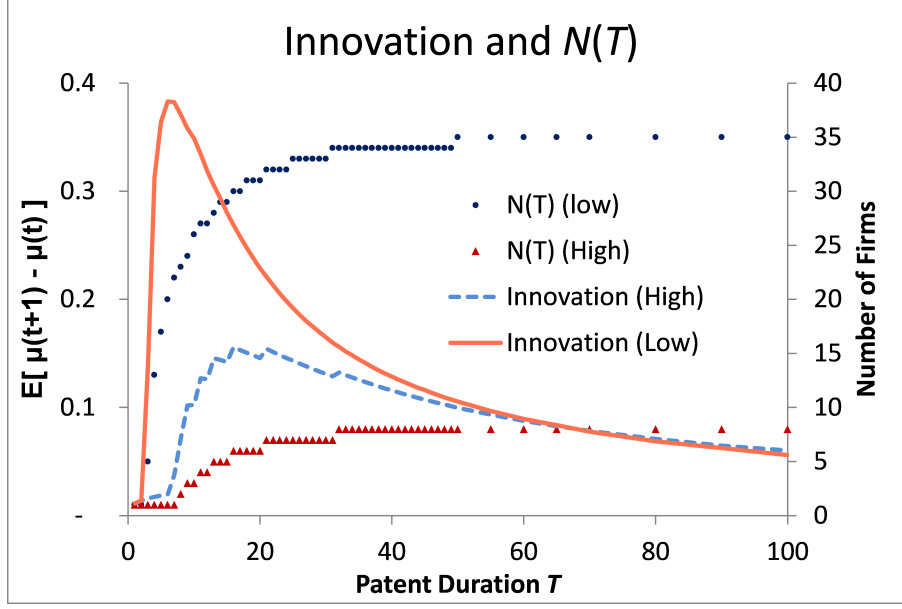


Figure 3.6: Innovation, under high and low fixed costs.

	T^*	$N(T^*)$	$\mathbf{E}[\mu_{t+1} - \mu_t]$	FC	R&D	Π	$\mathcal{T}_{\min}$	$N(\mathcal{T}_{\min})$
$x^e = .45$ (low)	6	20	.383	9.0	2.926	.29	3	5
$x^e = 1.35$ (high)	21	7	.154	9.45	2.773	.11	7	1

Table 3.3: T^* and $\mathcal{T}_{\min}$, under high and low fixed costs.

Figure 3.6 is intuitive; higher fixed cost industries optimally have more concentrated markets and lower rates of innovation. Qualitatively, our results are in agreement with the theoretical and empirical findings of Aghion et al. [2005]. Aghion et al. find an “inverted-U” relationship between market concentration and innovation.⁴³ In our model, free-entry means that increasing patent duration T is the same as increasing competition. We also find the inverted-U of innovation as a function of market competitiveness.

3.3.3 Comparative Statics

In this section I vary parameters and examine how T^* varies. Our baseline parametrization in this exercise is given in table 3.4. For each parameter that I vary, I consider three cases: high, medium

⁴³More specifically, they find an inverted-U relationship of innovation rates plotted against average Lerner-indices across firms within an industry-year. Their proxy for “innovation” is the number of patents granted; this is really problematic as the mapping from “patent” to “innovation” is far from perfect, but is common in the empirical literature in absence of a better observable.

Parameters		Functional Specifications		Entry (Fixed) Costs		
Δ	3	$\phi(I \dots)$	$\alpha I = .35I$		x^e	(x^f)
λ	1.1	$\eta(I \dots)$	.2	High	1.35	(.0675)
β	.95	$C(I)$	$500 \frac{I^4}{1-I}$	Medium	.90	(.0450)
c	.3	$U(\{X_t\})$	$\sum_{t=0}^{\infty} \beta^t \log(\sum_{\mu} \lambda^{\mu} X_t(\mu))$	Low	.45	(.0225)
$\bar{Y}$	5					

Table 3.4: Comparative Statics Baseline

and low fixed costs. The parameters that I examine are $\bar{\eta} \in [.1, .4]$, $\Delta \in [1, 4]$, $\lambda \in [1.05, 1.25]$ and $\phi(I) = \alpha I$ with $\alpha \in \{.2, .35, .5\}$ (see figures 3.7, 3.9, 3.10, 3.8, respectively).

I have set $\bar{Y} = 5$ and $\beta = .95$ which gives the present discounted value of all expenditures in the economy at 100. This is convenient as then the net present value of objects like R&D spending, fixed costs, etc, are normalized to be both magnitudes and percentages. In general, profits are about 0.6% of total expenditures or less, and I am comfortable ignoring them.

For each T , I solve the MPE for increasing N until $N + 1$ yields negative profits. When calculating MPEs I iterated on the value function until the sup norm between iterates was below $1.0e(-7)$. I then say $N(T) = N$. Given $N(T)$ I simulate the expected outcome under the MPE policy functions, using 10,000 sample paths of 500 periods each.⁴⁴ Given the simulations we can determine the welfare-maximizing T^* . We repeat this process every time we vary a parameter of the model. Solving for value functions takes anywhere from 50 to 600 iterations and between 80 and 7,000 ms. The simulations generally take between 1 and 2 ms per sample path draw. (Tables of the results of each parameter variation can be found in appendix B.7.)

My numerical exercises lead to the following ex-post policy conclusions. In general, patent length should be decreasing in “innovation technology”.⁴⁵ That is, higher $\bar{\eta}, \Delta, \lambda, \alpha$ all correspond to smaller optimal patent length. Smaller innovations correspond to lower monopolist rents (as the generic quality is almost as good). Therefore, it is harder to reward small incremental innovations

⁴⁴In some of the more jagged welfare-as-function-of- T cases, we needed to use upwards of 50,000 sample paths to compute the optimal T^* .

⁴⁵Slight non-monotonicities in T^* (and $N(T^*)$) arise due to the restriction that $N(T)$ be integral. This discreteness creates some weird situations where you are almost indifferent between having N firms and T patent duration versus $N + 1$ and $T + 1$. Ideally, the model wants something like $N + \frac{1}{2}$ firms, which I don't allow of course. Generally though, all of these curves are roughly monotonic in the parameters.

than large ones, so lengthier patents are optimal to incentivize participation. Smaller rates of innovation also correspond to higher market concentration as the fixed costs are constant but lower rates of innovation justify less firms operating. It is reasonable that T^* is decreasing in $\bar{\eta}$ as higher $\bar{\eta}$ indicates a higher probability of research exclusivity and less innovation. However, *ex-ante* I could easily justify T^* going in the other direction as well; we might wish to reward industries with a higher frequency of process innovation vs. incremental quality innovations with longer patents as the innovations are more fundamental. Similarly, when λ grows larger monopolist rents increase so we can understand why T^* is decreasing; *ex-ante* though one could also imagine higher- λ industries meriting longer patent lengths as individual innovations are worth more from a utility perspective so we would be willing to trade rents to get more innovation in these industries. In summary, I find that bigger quality and process innovations are more profitable, and hence the innovators do not need to be rewarded via longer patent length, in fact patent lengths decrease.⁴⁶

Interestingly, the optimal number of firms $N(T^*)$ is generally pretty flat in the parameters $\bar{\eta}$ and α ; innovation rate parameters. $N(T^*)$ is rapidly increasing in more utility/profit-related parameters like Δ and λ . When these are larger more firms enter as profits go up. It is interesting that $N(T^*)$ is not strongly increasing in α ; we might expect cheaper successes to recoup the higher fixed costs of more firms, but this does not appear to be the case. Industries that are more naturally successful holding investment fixed should have the same firm concentration.

As expected, higher fixed costs of operation or entry barriers (depends on interpretation) imply a smaller optimal number of firms and longer patent durations.⁴⁷ Similarly, raising α effectively lowers the cost of innovations across the board and I find that the lower costs of innovation necessitate shorter patent durations.⁴⁸

⁴⁶A weakness of our results is that innovation sizes are exogenous to all firm behavior, only probabilities of success. If I were to allow multiple jump sizes, i.e. $\mu' - \mu = \iota \in \{\bar{\iota}^0, \dots, \bar{\iota}^M\}$, I could potentially reverse this result. In the future it would be interesting to analyze a model with such a framework, where there was some probability distribution over innovation sizes that was a function of R&D decisions. In such a model I could analyze patentability requirements in addition to patent duration, something I am precluded from in our model.

⁴⁷This conclusion might flip with a dynamic entry/exit decision instead of time-0 entry decision.

⁴⁸As we discussed in section 3.1 our model is reasonably distant from much of the preceding literature on optimal patent design, so it is difficult to tie our findings back into earlier results.

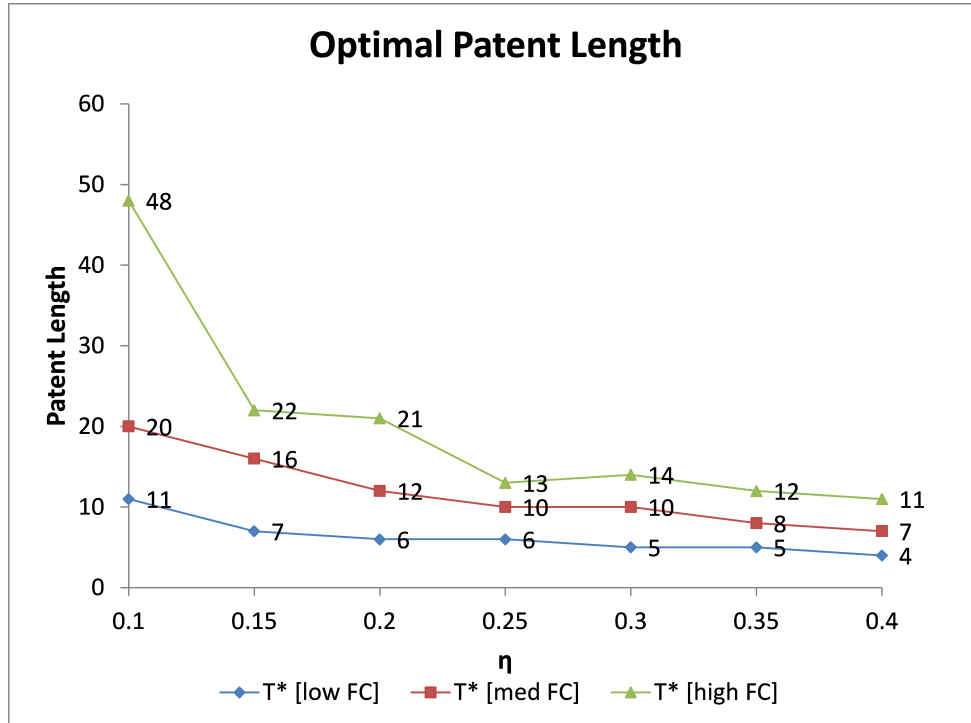


Figure 3.7: Varying $\bar{\eta}$

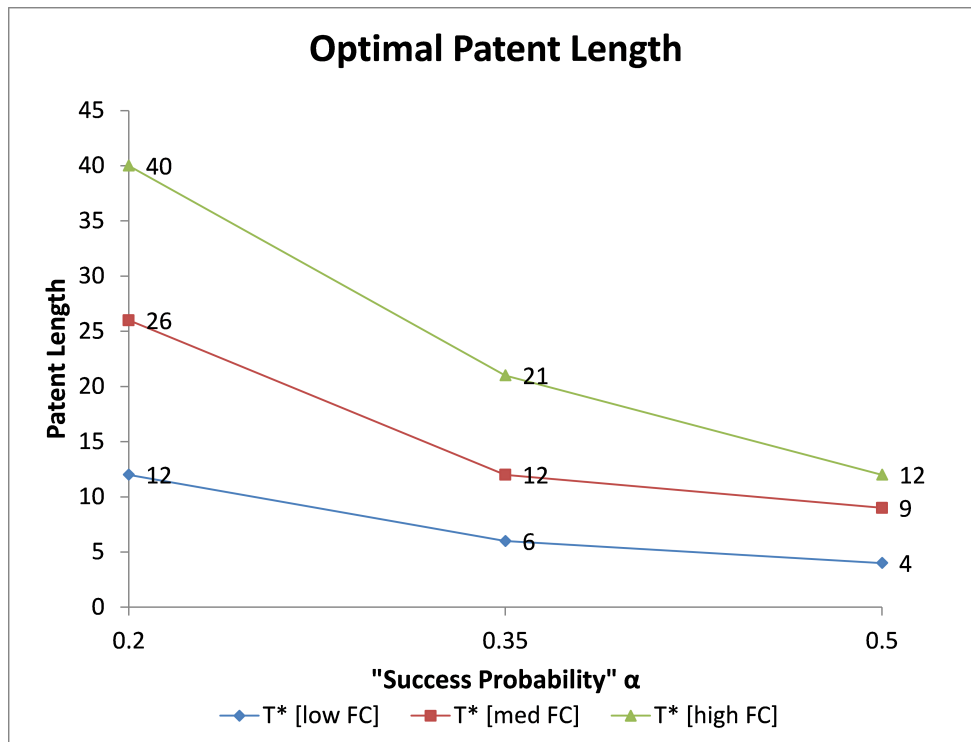


Figure 3.8: Varying α

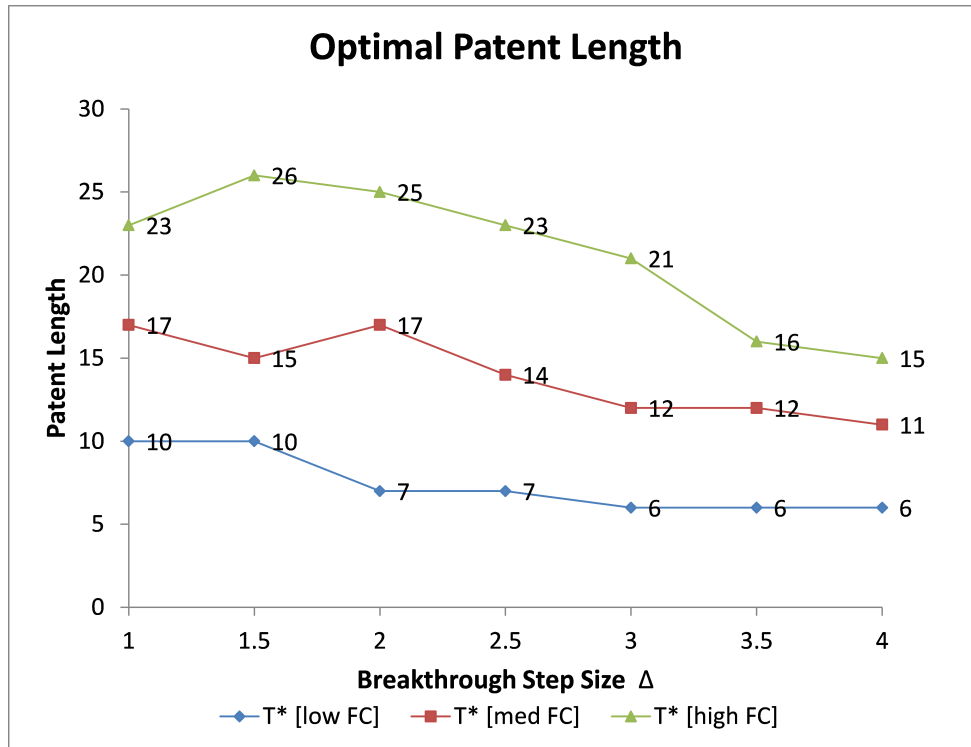


Figure 3.9: Varying Δ

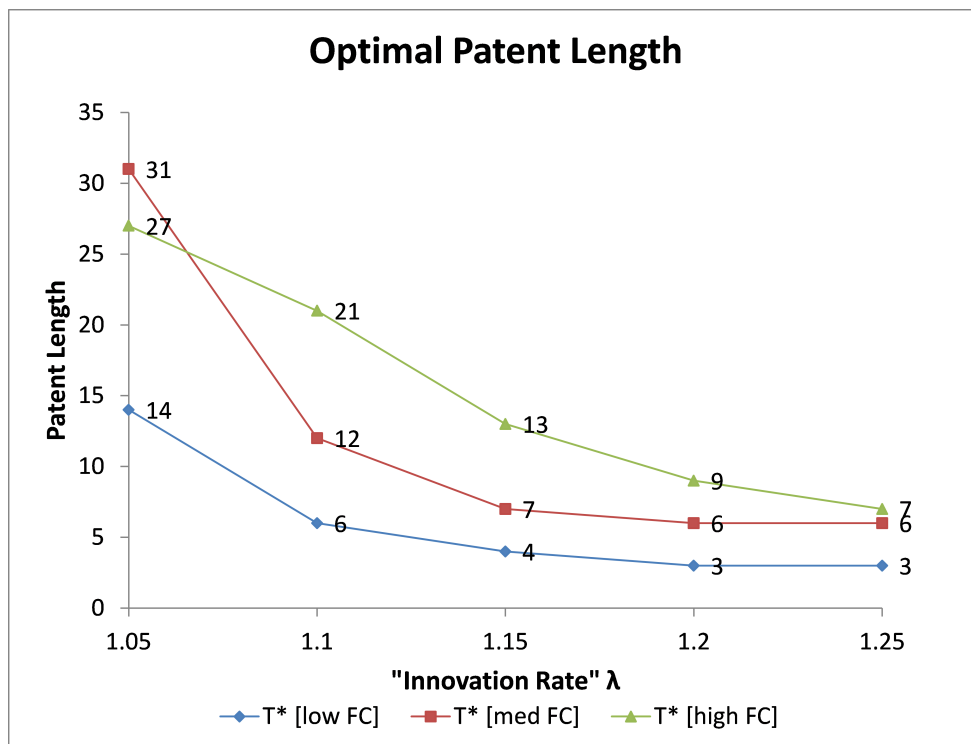


Figure 3.10: Varying λ

3.4 The Model Going Forward

3.4.1 Full Ericson-Pakes Entry/Exit Dynamics

The time-zero entry/exit assumption makes the state spaces and computations simpler, but there are tradeoffs. With time-zero entry/exit, increasing statutory patent duration T unambiguously increases the number of firms N who compete on innovation. (The number of firms asymptotes in my experience, but never decreases.) This need not be the case if I had period-by-period dynamic entry/exit choices. For example, if a firm is awarded a breakthrough patent (and hence research exclusivity), we might expect a rapid exodus of firms who would see no benefit in paying fixed costs for period after period as they sit idle. Similarly, we might expect a large influx of firms as the clock runs out on a breakthrough patent.⁴⁹

To add complete entry/exit dynamics, the number of firms N would join the state space; our states would be (χ, τ, ι, N) . In every period, incumbent firms would need to decide whether to remain or exit; and conditional on that an investment level. A pool of entrants would have to decide whether to enter, given the state.

I would also like to add more dynamic implications for investment. As the model stands, investment's only dynamic effect is on the change in the patent state. Unless you are successful and are awarded a patent, investment today is a complete loss *ex-post*. Following the literature established by Ericson and Pakes [1995a], I could posit a Markovian research productivity process ω_i that is specific to each firm. In full generality, we would have

$$\omega'_i = g(\omega_i, I_i) + \varepsilon'_i$$

where I_i is the firm's research intensity, g is some function that is increasing in I_i and ε'_i is some shock that is i.i.d. across time and firms. I can then allow the ϕ and η functions to depend on ω_i as well as I_i . Investment now productivity implications; we could see firms excluded under

⁴⁹When I say entry/exit, I am speaking of entering and exiting competition in a specific line of goods; for example hypertension medication. So our "firms" are really research divisions within firms.

breakthrough patents investing anyway to build up ω for when the patent expires. Suppose we restrict $\omega_i \in \{\bar{\omega}^0, \bar{\omega}^1, \dots, \bar{\omega}^K\}$.⁵⁰ The state now contains a vector ω where ω_k denotes the number of firms with productivity $\bar{\omega}^k$, and a variable ω^P which denotes the productivity level of the patent holder. The above two relaxations have the effect of *exploding* the state space. Stochastic methods might be necessary (see Powell [2011], Pakes and McGuire [2001]).

3.4.2 Empirical Analysis

This paper has provided a relatively parsimonious model of the effect of patent duration on innovation and consumer welfare in a dynamic oligopoly setting. Crucially, I emphasize how optimal patent length could vary across different industries. In the future, it is my hope to take this model to the data and do a cross-industry analysis of patent duration in the United States. The empirical literature on patents has been pretty inconclusive. I give a small (but in our view, representative) sample of the literature below. Hall and Ziedonis [2001] present a single industry-focused analysis including R&D data, while Lerner [2005] conducts a cross country analysis.

Hall and Ziedonis [2001] study the semiconductor industry from 1979-1995 which is a setting where cumulative innovation is an appropriate model. They find that while the increased patent strength over this period contributed to increased entry, they do not believe that firms used patents to recoup R&D investments. Instead, firms found it profitable to amass banks of patents for litigation battles. They find the variable $\frac{\text{patents issued}}{\text{R\&D spending}}$ increases over the time period they examine, but that patent “quality” (measured as average patent citations per patent) is also increasing. Lerner [2005] conducts a much broader analysis, examining patent applications across 60 countries across 150 years or so (using 177 patent policy changes as sources of identification).⁵¹ Lerner’s findings are ambiguous. He finds weak-to-no evidence that increased patent protection has corresponded to increased innovation.⁵² The author acknowledges several large problems with the analysis. It

⁵⁰It is common in the literature (see Doraszelski and Pakes [2007]) to assume that $K = 2$ and so there are high- and low-productivity states.

⁵¹This work is nicely summarized in Lerner [2009].

⁵²In some cases he find the opposite, that patent strengthening has led to decreased patent filings.

is hard to map statutory patent rules to effective patent protection since effective patent policy crucially depends on the rulings of individual USPTO employees as well as judges in subsequent litigation. Secondly, there are no good observables for the rate of innovation, forcing empiricists to use noisy, endogeneity-ridden, proxies.⁵³

Boldrin and Levine [2012] that a major issue with *any* empirical patent analysis is that “innovation” is such an abstract and unobservable process, it is hard to structurally map observables back into unobservables. The literature frequently uses patent applications/grants as a proxy for the innovation rate, which raises immediate endogeneity issues. Whether or not a firm seeks a patent (which in practice involves costly applications and litigation) will depend on the patent protection strength, regardless of the underlying innovation process. Strengthening patent protection could conceivably lead to increased patent filings without any change in the rate of the underlying innovations. A second difficulty is the reluctance of firms to divulge R&D data. As a result, empirical work on innovation has tended to focus on industry-wide aggregates rather than firm dynamics.

The model that I have discussed in this paper (along with the heterogeneity relaxations we’ve added in section 3.4.1) could provide a good starting point for the analysis of firm dynamics with respect to innovation. Roughly, the idea would be to do an Olley-Pakes-type analysis (Olley and Pakes [1992]). Treating innovative efficiency ω_{it} as the unobservable, we could try to estimate a patent production function using patents issues as the observable. For example, we could try to estimate a function

$$P_{it} = \exp \{ \omega_{it} + \xi_{it} \} f(X_{it}, I_{i,t-1})$$

where P_{it} is patents earned in that period, X_{it} is a vector of time- t input choices (research laborers, etc.), ξ_{it} is some shock unobservable to the firm and f is some function.⁵⁴ The objective would be to estimate parameters in our own model regarding the cost and innovation functions (with

⁵³Boldrin and Levine [2010b] (Ch. 8) provide a survey of two-dozen empirical papers on R&D and patents. In general, empirical results are pretty consistent with the findings of Lerner [2005], which is why I omit them.

⁵⁴Since we do not conduct this exercise in *this* paper, there is no sense in going into too much detail here. The methods of Olley and Pakes [1992] have prompted an extensive literature on dynamic production function estimation. Suffice it to say that down that path lies madness.

some mapping between $\{P_{it}\}_{i,t}$ and $\{\mu_t\}_t$). Given these estimates, we could run industry-specific counterfactuals and give quantitative policy recommendations about optimal patent duration.

3.5 Conclusion

I have constructed a model of quality-ladder innovation with patent protection through the duration instrument in a dynamic oligopoly setting. The paper makes a point of calculating endogenous market concentrations as a function of patent policies and primitives. I construct a model of innovation in a dynamic oligopoly with patent protection. The innovation process we construct distinguishes between product quality innovations (“incremental innovations”) and technology/process innovations (“breakthroughs”), which have different implications for patent enforcement. In the paper, the solution concept is symmetric Markov-perfect equilibrium. I discuss how to compute equilibria and simulate equilibria outcomes. In my experiments I varied patent duration T and computed optimal patent duration T^* and market concentration $N(T^*)$. I discuss how T^* varies in the model primitives. Product lines with high entry/fixed costs should have lengthy patent durations and small firm concentrations. Product lines characterized by rapid innovation rates or high probabilities of breakthroughs generally warrant shorter patent duration and lower market concentration.

Bibliography

2014 NPE Cost Report. Technical report, RPX Corporation, 2014. URL <http://chipsnetwork.org/wp-content/uploads/2012/09/2014-NPE-Cost-Report.pdf>.

Jaap Abbring, Jeffrey Campbell, and Nan Yang. Simple Markov-Perfect Industry Dynamics. Technical report, 2012.

J.H. Abbring, J.R. Campbell, J. Tilly, and N. Yang. Very Simple Markov-Perfect Industry Dynamics. Technical report, 2014.

Jose Miguel Abito. How much can we identify from repeated games? MPRA Paper 66378, University Library of Munich, Germany, August 2015. URL <https://ideas.repec.org/p/pra/MPRA/66378.html>.

Dilip Abreu. Extremal equilibria of oligopolistic supergames. *Journal of Economic Theory*, 39(1): 191–225, 1986. URL <http://EconPapers.repec.org/RePEc:eee:jetho:v:39:y:1986:i:1:p:191-225>.

Dilip Abreu. On the theory of infinitely repeated games with discounting. *Econometrica*, 56(2): 383–96, 1988. URL <http://EconPapers.repec.org/RePEc:ecm:emetrp:v:56:y:1988:i:2:p:383-96>.

Dilip Abreu and Yuliy Sannikov. An algorithm for two-player repeated games with perfect monitoring. *Theoretical Economics*, 9(2), 2014. URL <http://EconPapers.repec.org/RePEc:the:publsh:1302>.

- Dilip Abreu, David Pearce, and Ennio Stacchetti. Optimal cartel equilibria with imperfect monitoring. *Journal of Economic Theory*, 39(1):251–269, 1986. URL <http://EconPapers.repec.org/RePEc:eee:jetheo:v:39:y:1986:i:1:p:251-269>.
- Dilip Abreu, David Pearce, and Ennio Stacchetti. Toward a theory of discounted repeated games with imperfect monitoring. *Econometrica*, 58(5):1041–1063, 1990. ISSN 00129682, 14680262. URL <http://www.jstor.org/stable/2938299>.
- Dilip Abreu, Ben Brooks, and Yuliy Sannikov. A pencil-sharpening algorithm for two player stochastic games with perfect monitoring. Technical report, 2016. URL http://www.benjaminbrooks.net/downloads/abs_stochasticgames.pdf.
- Philippe Aghion, Nick Bloom, Richard Blundell, Rachel Griffith, and Peter Howitt. Competition and innovation: An inverted-u relationship. *The Quarterly Journal of Economics*, 120(2):701–728, May 2005.
- Paul McGuire Ariel Pakes. Computing markov-perfect nash equilibria: Numerical implications of a dynamic differentiated product model. *The RAND Journal of Economics*, 25(4):555–589, 1994. ISSN 07416261. URL <http://www.jstor.org/stable/2555975>.
- Ashish Arora, Marco Ceccagnoli, and Wesley M. Cohen. R&D and the patent premium. *International Journal of Industrial Organization*, pages 1153–1179, 2008.
- Susan Athey and Kyle Bagwell. Optimal collusion with private information. *RAND Journal of Economics*, 32(3):428–65, 2001. URL <http://EconPapers.repec.org/RePEc:rje:randje:v:32:y:2001:i:3:p:428-65>.
- Andrew Atkeson and Ariel T. Burstein. Aggregate Implications of Innovation Policy. NBER Working Papers 17493, National Bureau of Economic Research, Inc, October 2011. URL <http://ideas.repec.org/p/nbr/nberwo/17493.html>.

- Roberto Baldwin. Apple patents the macbook air's wedge design. *Wired*, June 2012. URL <http://www.wired.com/gadgetlab/2012/06/apple-patents-the-macbook-air-s-wedge-design-bad-news-for-ultrabook-makers/>.
- Paul M. Barret. Apple's war on android. *Bloomberg Business Week*, March 2012. URL <http://www.businessweek.com/articles/2012-03-29/apple-s-war-on-android>.
- K Berg and M Kitti. Fractal geometry of equilibrium payoffs in discounted supergames. *fractals* (to appear), 2014.
- David Besanko, Ulrich Doraszelski, Yaroslav Kryukov, and Mark Satterthwaite. Learning-by-doing, organizational forgetting, and industry dynamics. *Econometrica*, 78(2):453–508, 2010. ISSN 1468-0262. doi: 10.3982/ECTA6994. URL <http://dx.doi.org/10.3982/ECTA6994>.
- James Bessen. Estimates of patent rents from firm market value. *Research Policy*, 38(10):1604–1616, December 2009. URL <http://ideas.repec.org/a/eee/respol/v38y2009i10p1604-1616.html>.
- James Bessen and Eric Maskin. Sequential innovation, patents, and imitation. *RAND Journal of Economics*, 40(4):611–635, 2009.
- James Bessen and Michael J Meurer. The direct costs from npe disputes. *Cornell Law Review*, 99(2), January 2014.
- James Bessen, Jennifer L. Ford, and Michael J. Meurer. The Private and Social Costs of Patent Trolls. Technical report, 2011.
- Michele Boldrin and David K Levine. Quality ladders, competition and endogenous growth. Technical report, 2010a.
- Michele Boldrin and David K. Levine. *Against Intellectual Monopoly*. Number 9780521127264 in Cambridge Books. Cambridge University Press, December 2010b.

- Michele Boldrin and David K. Levine. The case against patents. Technical report, 2012.
- Ron N. Borkovsky, Ulrich Doraszelski, and Yaroslav Kryukov. A dynamic quality ladder model with entry and exit: Exploring the equilibrium correspondence using the homotopy method. *Quantitative Marketing and Economics*, 10(2):197–229, 2012. ISSN 1570-7156. doi: 10.1007/s11129-011-9113-4. URL <http://dx.doi.org/10.1007/s11129-011-9113-4>.
- Eric Budish, Benjamin Roin, and Heidi Williams. Do fixed patent terms distort innovation? Evidence from cancer clinical trials. Discussion Papers 13-001, Stanford Institute for Economic Policy Research, September 2013. URL <http://ideas.repec.org/p/sip/dpaper/13-001.html>.
- Owen Byrd and Brian Howard. 2013 patent litigation in review. Technical report, Lex Machina, 2014. URL [https://www.law.berkeley.edu/files/2013_Patent_Litigation_Year_in_Review_Full_Report_\(MLex_Machina\).pdf](https://www.law.berkeley.edu/files/2013_Patent_Litigation_Year_in_Review_Full_Report_(MLex_Machina).pdf).
- Howard F. Chang. Patent scope, antitrust policy, and cumulative innovation. *RAND Journal of Economics*, 26(1):34–57, 1995.
- James Conklin. Computing Value Correspondences for Repeated Games with State Variables. Technical report, 1996.
- Mark B Cronshaw. Algorithms for finding repeated game equilibria. *Computational Economics*, 10(2):139–168, 1997.
- Eric V. Denardo. Contraction mappings in the theory underlying dynamic programming. *SIAM Review*, 9(2), 1967.
- Vincenzo Denicolo. Two-stage patent races and patent policy. *RAND Journal of Economics*, 31(3):488–501, 2000.
- Ulrich Doraszelski. Innovations, improvements, and the optimal adoption of new technologies. *Journal of Economic Dynamics and Control*, 28(7):1461 – 1480, 2004. ISSN 0165-1889. doi:

[http://dx.doi.org/10.1016/S0165-1889\(03\)00112-X](http://dx.doi.org/10.1016/S0165-1889(03)00112-X). URL <http://www.sciencedirect.com/science/article/pii/S016518890300112X>.

Ulrich Doraszelski and Juan F. Escobar. A theory of regular markov perfect equilibria in dynamic stochastic games: Genericity, stability, and purification. *Theoretical Economics*, 5(3):369–402, 2010. ISSN 1555-7561. doi: 10.3982/TE632. URL <http://dx.doi.org/10.3982/TE632>.

Ulrich Doraszelski and Jordi Jaumandreu. R&d and productivity: Estimating endogenous productivity. *The Review of Economic Studies*, 2013. doi: 10.1093/restud/rdt011. URL <http://restud.oxfordjournals.org/content/early/2013/02/23/restud.rdt011.abstract>.

Ulrich Doraszelski and Kenneth L. Judd. Avoiding the curse of dimensionality in dynamic stochastic games. *Quantitative Economics*, 3(1):53–93, 03 2012.

Ulrich Doraszelski and Ariel Pakes. *A Framework for Applied Dynamic Analysis in IO*, volume 3 of *Handbook of Industrial Organization*, chapter 30, pages 1887–1966. Elsevier, June 2007.

Ulrich Doraszelski and Mark Satterthwaite. Computable markov-perfect industry dynamics: Existence, purification, and multiplicity. Cepr discussion papers, C.E.P.R. Discussion Papers, March 2007.

Emmanuel Duguet and Claire Lelarge. Does patenting increase the private incentives to innovate? A microeconomic analysis. Cahiers de la Maison des Sciences Economiques v04096, Université Pantheon-Sorbonne (Paris 1), September 2004. URL <http://ideas.repec.org/p/mse/wpsorb/v04096.html>.

Charles Duhigg and Steve Lohr. The patent, used as a sword. *New York Times*, October 2012. URL <http://www.nytimes.com/2012/10/08/technology/patent-wars-among-tech-giants-can-stifle-competition.html?pagewanted=all&r=0>.

- Robert H. Porter Edward J. Green. Noncooperative collusion under imperfect price information. *Econometrica*, 52(1):87–100, 1984. ISSN 00129682, 14680262. URL <http://www.jstor.org/stable/1911462>.
- Richard Ericson and Ariel Pakes. Markov-perfect industry dynamics: A framework for empirical work. *Review of Economic Studies*, 62(1):53–82, January 1995a.
- Richard Ericson and Ariel Pakes. Markov-perfect industry dynamics: A framework for empirical work. *The Review of Economic Studies*, 62(1):53–82, 1995b.
- Executive Office of the President. Patent assertion and u.s. innovation. June 2014. URL https://www.whitehouse.gov/sites/default/files/docs/patent_report.pdf.
- Chaim Fershtman and Ariel Pakes. A dynamic oligopoly with collusion and price wars. *The RAND Journal of Economics*, 31(2):207–236, 2000. ISSN 07416261. URL <http://www.jstor.org/stable/2601038>.
- Chaim Fershtman and Ariel Pakes. Dynamic Games with Asymmetric Information: A Framework for Empirical Work. *The Quarterly Journal of Economics*, 127(4):1611–1661, 2012a. URL <https://ideas.repec.org/a/oup/qjecon/v127y2012i4p1611-1661.html>.
- Chaim Fershtman and Ariel Pakes. Dynamic games with asymmetric information: A framework for empirical work. *The Quarterly Journal of Economics*, page qjs025, 2012b.
- Alberto Galasso and Mark Schankerman. Patents and cumulative innovation: Causal evidence from the courts. Working Paper 20269, National Bureau of Economic Research, July 2014. URL <http://www.nber.org/papers/w20269>.
- Nancy T. Gallini. Patent policy and costly imitation. *RAND Journal of Economics*, 23(1):52–63, Spring 1992.
- Richard Gilbert and Carl Shapiro. Optimal patent length and breadth. *The RAND Journal of Economics*, 21:106–112, 1990.

- Ronald L. Goettler and Brett R. Gordon. Does AMD Spur Intel to Innovate More? *Journal of Political Economy*, 119(6):1141 – 1200, 2011. URL <http://ideas.repec.org/a/ucp/jpolec/doi10.1086-664615.html>.
- Ronald L. Goettler and Brett R. Gordon. Competition and product innovation in dynamic oligopoly. *Quantitative Marketing and Economics*, 12(1):1–42, 2014. ISSN 1570-7156. doi: 10.1007/s11129-013-9142-2. URL <http://dx.doi.org/10.1007/s11129-013-9142-2>.
- Susanne Goldlücke and Sebastian Kranz. Infinitely repeated games with public monitoring and monetary transfers. *Journal of Economic Theory*, 147(3):1191–1221, 2012.
- Jerry R. Green and Suzanne Scotchmer. On the division of profit in sequential innovation. *RAND Journal of Economics*, 26(1):20–33, Spring 1995.
- Gene M Grossman and Elhanan Helpman. Quality ladders in the theory of growth. *Review of Economic Studies*, 58(1):43–61, January 1991.
- Gene M. Grossman and Elhanan Helpman. *Innovation and Growth in the Global Economy*, volume 1 of *MIT Press Books*. The MIT Press, 1993.
- Inc. Gurobi Optimization. Gurobi optimizer reference manual, 2015. URL <http://www.gurobi.com>.
- Bronwyn H Hall and Rosemarie Ham Ziedonis. The patent paradox revisited: An empirical study of patenting in the u.s. semiconductor industry, 1979-1995. *RAND Journal of Economics*, 32(1):101–28, Spring 2001.
- Julian Hatter. Obama urges patent reform. The Hill, October 2014. URL <http://thehill.com/policy/technology/220371-obama-urges-patent-reform>.
- Hugo Hopenhayn and Matthew Mitchell. Optimal patent policy with recurrent innovators. Technical report, 2010.

- Hugo Hopenhayn, Gerard Llobet, and Matthew Mitchell. Rewarding sequential innovators: Prizes, patents, and buyouts. *Journal of Political Economy*, 114(6):1041–1068, December 2006.
- Hugo A Hopenhayn and Matthew F Mitchell. Innovation variety and patent breadth. *RAND Journal of Economics*, 32(1):152–66, Spring 2001.
- Andrew W. Horowitz and Edwin L.C. Lai. Patent length and the rate of innovation. *International Economic Review*, 37(4):785–801, 1996.
- Brian Howard. 2014 patent litigation in review. Technical report, Lex Machina, 2015. URL <http://pages.lexmachina.com/rs/lexmachina/images/2014%20Patent%20Litigation%20Report.pdf>.
- Robert M. Hunt. Patentability, industry structure, and innovation. *Journal of Industrial Economics*, 52(3):401–425, 09 2004.
- Mitsuru Igami. Estimating the innovator’s dilemma: Structural analysis of creative destruction in the hard disk drive industry. 2011.
- Bessen James. Evaluating the Economic Performance of Property Systems. *Review of Law & Economics*, 5(3):1037–1061, December 2009. URL <http://ideas.repec.org/a/bpj/rlecon/v5y2009i3n4.html>.
- Kenneth Judd and Sevin Yeltekin. Computing equilibria of dynamic games. GSIA Working Papers 2011-E20, Carnegie Mellon University, Tepper School of Business. URL <http://EconPapers.repec.org/RePEc:cmu:gsiawp:-962041374>.
- Kenneth L. Judd, Sevin Yeltekin, and James Conklin. Computing Supergame Equilibria. *Econometrica*, 71(4):1239–1254, 07 2003. URL <https://ideas.repec.org/a/ecm/emetrp/v71y2003i4p1239-1254.html>.
- Richard M. Katzwer. Optimal patent duration in a dynamic oligopoly. 2012.

- Paul Klemperer. How broad should the scope of patent protection be? *RAND Journal of Economics*, 21(1):113–130, 1990.
- T.J. Klette. Should Monopolized Industries Have Infite Patent Life? Technical report, 1990.
- Josh Lerner. 150 years of patent office practice. *American Law and Economics Review*, 7(1): 112–143, 2005.
- Josh Lerner. The empirical impact of intellectual property rights on innovation: Puzzles and clues. *American Economic Review*, 99(2):343–48, May 2009.
- Jonathan Levin. Relational incentive contracts. *American Economic Review*, 93(3):835–857, 2003. URL <http://EconPapers.repec.org/RePEc:aea:aecrev:v:93:y:2003:i:3:p:835-857>.
- Lars Ljungqvist and Thomas J. Sargent. *Recursive Macroeconomic Theory, 2nd Edition*, volume 1. The MIT Press, 2 edition, 2004. URL <http://EconPapers.repec.org/RePEc:mtp:titles:026212274x>.
- Steve Lohr. U.s. disavows patent at center of apple-samsung dispute. *New York Times*, October 2012. URL <http://www.nytimes.com/2012/10/24/technology/us-disavows-patent-at-center-of-apple-samsung-dispute.html>.
- George Mailath and Larry Samuelson. *Repeated Games and Reputations: Long-Run Relationships*. Oxford University Press, 2006. URL <http://EconPapers.repec.org/RePEc:exp:obooks:9780195300796>.
- Eric Maskin. Patents on software: A nobel laureate’s view. *New York Times*, October 2012. URL <http://www.nytimes.com/2012/10/15/opinion/patents-on-software-a-nobel-laureates-view.html>.
- Eric Maskin and Jean Tirole. A theory of dynamic oligopoly, i: Overview and quantity competition with large fixed costs. *Econometrica*, 56(3):549–69, May 1988.

Eric Maskin and Jean Tirole. Markov perfect equilibrium: I. observable actions. *Journal of Economic Theory*, 100(2):191–219, October 2001.

Ben Mermelstein, Volker Nocke, Mark A. Satterthwaite, and Michael D. Whinston. Internal versus External Growth in Industries with Scale Economies: A Computational Model of Optimal Merger Policy. NBER Working Papers 20051, National Bureau of Economic Research, Inc, April 2014. URL <https://ideas.repec.org/p/nbr/nberwo/20051.html>.

Joe Mullin. The year in patent litigation: More trolling, more texas. *Arstechnica.com*, May 2014. URL <http://arstechnica.com/tech-policy/2014/05/the-year-in-patent-litigation-more-trolling-more-texas/>.

William D. Nordhaus. The optimal life of a patent. Cowles foundation discussion papers, Cowles Foundation for Research in Economics, Yale University, November 1967.

Ted O’Donoghue. A patentability requirement for sequential innovation. *RAND Journal of Economics*, 29(4):654–679, Winter 1998.

Ted O’Donoghue, Suzanne Scotchmer, and Jacques-Francois Thisse. Patent breadth, patent life, and the pace of technological progress. *Journal of Economics & Management Strategy*, 7(1): 1–32, 03 1998.

James F. Oehmke and Christopher A. Wolf. Measuring concentration in the biotechnology r&d industry: Adjusting for interfirm transfer of genetic materials. *The Journal of Agrobiotechnology Management & Economics*, 6(3):134–140, 2003.

U.S. Congress Office of Technology Assessment. Pharmaceutical r&d: Cost, risks and rewards. 1993.

Office of the Press Secretary. Executive actions: Answering the president’s call to strengthen our patent system and foster innovation, February

2014. URL <https://www.whitehouse.gov/the-press-office/2014/02/20/fact-sheet-executive-actions-answering-president-s-call-strengthen-our-p>.
- G. Steven Olley and Ariel Pakes. The dynamics of productivity in the telecommunications equipment industry. Nber working papers, National Bureau of Economic Research, Inc, January 1992.
- Ariel Pakes. Patents as options: Some estimates of the value of holding european patent stocks. Nber working papers, National Bureau of Economic Research, Inc, November 1986.
- Ariel Pakes and Paul McGuire. Computing markov-perfect nash equilibria: Numerical implications of a dynamic differentiated product model. *RAND Journal of Economics*, 25(4):555–589, Winter 1994.
- Ariel Pakes and Paul McGuire. Stochastic algorithms, symmetric markov perfect equilibrium, and the ‘curse’ of dimensionality. *Econometrica*, 69(5):1261–81, 2001.
- Robert H. Porter. A study of cartel stability: The joint executive committee, 1880-1886. *The Bell Journal of Economics*, 14(2):301–314, 1983. ISSN 0361915X. URL <http://www.jstor.org/stable/3003634>.
- Warren B. Powell. *Approximate Dynamic Programming: Solving the Curses of Dimensionality*. John Wiley and Sons, Cambridge, MA, 2nd edition, 2011.
- The Associated Press. Uk: Apple must apologize again over copycat claims. *The Associated Press*, November 2012. URL <http://www.usatoday.com/story/tech/2012/11/01/uk-apple-must-apologize-again-over-copycat-claims/1674915/>.
- Roy Radner. Repeated partnership games with imperfect monitoring and no discounting. *Review of Economic Studies*, 53(1):43–57, 1986.
- Katharine Rockett. Chapter 7: Property rights and invention. *Handbook of The Economics of Innovation*, 1, 2010.

- Yuliy Sannikov. Games with imperfectly observable actions in continuous time. *Econometrica*, 75 (5):1285–1329, 2007.
- F.M. Scherer. Chapter 12: Pharmaceutical innovation. *Handbook of The Economics of Innovation*, 1, 2010.
- Joseph A. Schumpeter. *Capitalism, socialism, and democracy*. Routledge, New York, 1994 (1942).
- Suzanne Scotchmer. Protecting early innovators: Should second-generation products be patentable? *RAND Journal of Economics*, 27(2):322–331, Summer 1996.
- Suzanne Scotchmer. *Innovation and Incentives*, volume 1 of *MIT Press Books*. The MIT Press, 2006.
- Ilya Segal and Michael D. Whinston. Antitrust in innovative industries. *American Economic Review*, 97(5):1703–1730, 2007. doi: 10.1257/aer.97.5.1703. URL <http://www.aeaweb.org/articles.php?doi=10.1257/aer.97.5.1703>.
- Carl Shapiro. Patent reform: Aligning reward and contribution. In *Innovation Policy and the Economy*, Volume 8, NBER Chapters, pages 111–156. National Bureau of Economic Research, Inc, August 2008.
- Greg Stohr. Monsanto seed patent case gets u.s. supreme court review. *Bloomberg Business Week*, October 2012. URL <http://www.businessweek.com/news/2012-10-05/monsanto-seed-patent-case-gets-u-dot-s-dot-supreme-court-review>.
- Nancy L. Stokey, Robert E. Lucas, Jr., and Edward C. Prescott. *Recursive Methods in Economic Dynamics*. Harvard University Press, Cambridge, MA, 1989.
- Alex Tabarrok. Patent policy on the back of a napkin. *marginalrevolution.com*, September 2012. URL <http://marginalrevolution.com/marginalrevolution/2012/09/patent-theory-on-the-back-of-a-napkin.html>.

- Pankaj Tandon. Optimal patents with compulsory licensing. *Journal of Political Economy*, 90(3): 470–86, June 1982.
- Jean Tirole. *The Theory of Industrial Organization*, volume 1 of *MIT Press Books*. The MIT Press, 1988.
- Ward Whitt. Representation and approximation of noncooperative and sequential games. *SIAM Journal of Control and Optimization*, 18(1), January 1980.
- Heidi L. Williams. Intellectual Property Rights and Innovation: Evidence from the Human Genome. *Journal of Political Economy*, 121(1):1 – 27, 2013. URL <http://ideas.repec.org/a/ucp/jpolec/doi10.1086-669706.html>.
- Sevin Yeltekin, Yongyang Cai, and Kenneth L Judd. Computing equilibria of dynamic games. 2016.
- Ali Yurukoglu. Medicare reimbursements and shortages of sterile injectable pharmaceuticals. *NBER Working Paper No. 17987*, April 2012.

Appendix A

Appendices for Ch. 1

A.1 The Static Model in 1.4.2.1

I present here the static model that was used above for static-dynamic model prediction comparisons. It was constructed to be as close to the dynamic model as possible. Firms enter, invest, innovate, litigate, and earn profits. Consumers earn consumer surplus from innovations. Firms are given a T -period statutory patent with lagging breadth $\delta \in \{1, \dots, \Delta\}$, but every period with probability $\zeta \in [0, 1)$, the patent is invalidated or rendered obsolete.

Market profits The consumers behave identically to those in the dynamic model. A firm with the right-to-practice quality gap δ will price at $p^*(\delta) = \bar{c} + \delta z^*$.¹ Period market profits to the patentee t years into a patent are then $\pi_t(\delta) = M_{T-t} \delta z^* [1 - F(z^*)]$. Therefore, with a discount rate of β , the expected discounted market profits to a firm looking forward at T periods of statutory right-to-practice is

$$\Pi(\delta, T) = \sum_{t=1}^T \left[\beta (1 - \zeta_\delta) \right]^t \pi_t(\delta).$$

¹Recall, if F_λ is the taste-for-quality CDF then z^* solves $\max_z z(1 - F_\lambda(z))$.

Patent assignment and litigation Suppose the firms $i = 1, 2, \dots, N$ each draw a quality innovation δ_i . Let $\delta^{\max} = \max_i \delta_i$ and $n^{\max}$ be the number of firms who achieve $\delta^{\max}$. If $\delta^{\max} = 0$ (no innovation), no patent is assigned. Otherwise, if $n^{\max} = 1$, that firm is assigned a patent of quality $\delta^{\max}$ for T periods. If $n^{\max} > 1$ and $\delta^{\max} > 0$, I have multiple firms simultaneously “winning” the patent race, and the right to practice is resolved through the courts. Each firm $i = 1, \dots, n^{\max}$ chooses litigation spending ℓ_i . Firm i wins the right to practice with probability $\chi(\ell_i, \ell_{-i})$,² and solves the problem

$$\max_{\ell_i} \chi(\ell_i, \ell_{-i}) \Pi(\delta^{\max}, T) - \ell_i.$$

I assume the symmetric Nash equilibrium in this litigation subgame.

R&D investment Given the patent assignment rule above, each of the research firms $i = 1, \dots, N$ chooses an R&D level x_i to maximize expected discounted profits,

$$\mathbf{E} \left[- (1 - \bar{a}^x) x_i - \ell_i + \mathbf{1}_{\{i \text{ has right to practice } \delta\}} \Pi(\delta, T) \mid N, x_{-i}, \ell_{-i} \right].$$

Given investments (x_i, x_{-i}) , firm i draws innovation $\delta_i \in \{0, 1, \dots, \Delta\}$ with probability $\eta(\delta_i \mid x_i, x_{-i})$.

I assume the symmetric Nash equilibrium in the investment game.

Entry At time-zero, N firms enter and pay an entry fixed cost $\tilde{x}^e$.³ Assume that firms enter to satisfy an approximate zero-profit condition: N is the largest number of firms for which profits are larger than $\tilde{x}^e$.

²Where χ is the same as in the dynamic model in the case of no-existing patentee.

³ $\tilde{x}^e = (1 - \bar{a}^e) \bar{x}^e + (1 - \bar{a}^f) \bar{x}^f - \beta \bar{x}^s$, where the right-hand-side comes are parameters from the dynamic model.

Welfare In this model, welfare is taken to be

$$W_{\text{static}} = \mathbf{E} \left[\begin{aligned} & (\mathbf{E}[\lambda_i] \delta - \bar{c}) \sum_{t=1}^{\infty} \beta^t M_t - \bar{a}^x \sum_{i=1}^N x_i - N \left(\bar{a}^e \bar{x}^e + \bar{a}^f \bar{x}^f \right) - \kappa^\ell \sum_{i=1}^N \ell_i \\ & - \delta \sum_{t=1}^T (\beta (1 - \zeta))^t M_t [\mathbf{E}[\lambda_i \mid \lambda_i \leq z^*] F_\lambda(z^*) + z^* (1 - F_\lambda(z^*))] \end{aligned} \right]$$

A.2 Baseline Functional Specification in Section 1.4.1

Parameter		Low Cost	High Cost
Discount Rate	β	.96	“
Entry Cost	$\bar{x}^e$	2.5	“
Scrap Value	$\bar{x}^s$	.75	“
Fixed R&D Cost	$\bar{x}^f$	.30	.45
Pr of $\delta = 0$	$\eta(0 \mid x_i)$	$1 - \frac{.4x_i^{4/5}}{5+x_i^{4/5}}$	$1 - \frac{.267x_i^{4/5}}{5+x_i^{4/5}}$
Pr of $\delta = 1$	$\eta(1 \mid x_i)$	$\frac{.4x_i^{4/5}}{5+x_i^{4/5}} \cdot .625$	$\frac{.267x_i^{4/5}}{5+x_i^{4/5}} \cdot .625$
Pr of $\delta = 2$	$\eta(2 \mid x_i)$	$\frac{.4x_i^{4/5}}{5+x_i^{4/5}} \cdot .3$	$\frac{.267x_i^{4/5}}{5+x_i^{4/5}} \cdot .3$
Pr of $\delta = 3$	$\eta(3 \mid x_i)$	$\frac{.4x_i^{4/5}}{5+x_i^{4/5}} \cdot .075$	$\frac{.267x_i^{4/5}}{5+x_i^{4/5}} \cdot .075$
Market Size	M	12	“
Quality Taste Distribution	$F_\lambda(z)$	$1 - \exp\{.5z\}$	“
Unit Cost of Production	$\bar{c}$	.15	“
Patent Invalidation Risk	$\zeta(T - \tau)$	.005	“

Table A.1: Primitives

Parameter	High- and Low-Cost	
Statutory Duration	T	12
Renewal Fee	$a(\mathbf{s}_t)$	0
Variable R&D Cost Subsidy	$\bar{a}^x$	0
Fixed R&D Cost Subsidy	$\bar{a}^f$	0
Entry Cost Subsidy	a^e	0
Pr i wins litigation (no patentee)		$\frac{\ell_i^{A/5}}{\sum_j \ell_j^{A/5}}$
Pr P wins litigation (active patentee)		$.25 + \frac{.75(1-.1(\delta^C - \delta^P))\ell_P^{A/5}}{\sum_{j \neq P} \ell_j^{A/5} + (1-.1(\delta^C - \delta^P))\ell_P^{A/5}}$
Pr $\neg P$ wins litigation (active patentee)		$\frac{.75\ell_i^{A/5}}{\sum_{j \neq P} \ell_j^{A/5} + (1-.1(\delta^C - \delta^P))\ell_P^{A/5}}$

Table A.2: Government Policy

Appendix B

Appendices for Ch. 3

B.1 Utility Specifications in 3

B.1.1 Utility

B.1.1.1 Utility Specification #1

The standard literature on quality ladders uses the utility specification from Grossman and Helpman [1991],¹

$$U \equiv \mathbf{E}_0 \sum_{t=0}^{\infty} \beta^t \log \left(\sum_{\mu=0}^{\mu_t} \lambda^{\mu} X_t(\mu) \right), \quad \lambda > 1$$

where the consumer solves

$$\begin{aligned} & \max_{\{X_t(\cdot)\}_{t=0}^{\infty}} U \\ & \text{s.t.} \quad \sum_{t=0}^{\infty} \beta^t \sum_{\mu=0}^{\mu_t} p_t(\mu) X_t(\mu) \leq \frac{1}{1-\beta} \bar{Y} \end{aligned}$$

¹In Grossman and Helpman [1991] there is actually a continuum of goods indexed by $g \in [0, 1]$ with evolving qualities μ_t^g , and (modified for discrete time)

$$U \equiv \mathbf{E}_0 \sum_{t=0}^{\infty} \beta^t \int_0^1 \log \left(\sum_{\mu=0}^{\mu_t^g} \lambda^{\mu} X_t^g(\mu) \right) dg.$$

where $p_t(\mu)$ is the time- t price of quality μ , and $\frac{1}{1-\beta}\bar{Y}$ is some exogenous lifetime income.² The household solves a two-step optimization:

1. **Static Optimization:** Suppose the household budgets expenditure $\mathcal{E}_t$ for period t . The household then solves

$$\begin{aligned} \max_{X_t(\cdot)} \quad & \sum_{\mu=0}^{\mu_t} \lambda^\mu X_t(\mu) \\ \text{s.t.} \quad & \sum_{\mu=0}^{\mu_t} p_t(\mu) X_t(\mu) \leq \mathcal{E}_t \end{aligned}$$

Then it is immediate that the household consumes only quality

$$\mu_t^* = \arg \max_{\mu \leq \mu_t} \{\lambda^\mu / p_t(\mu)\},$$

the level with the highest quality-to-price ratio. The household demands $X_t(\mu_t^*) = \mathcal{E}_t / p_t(\mu_t^*)$.

2. **Dynamic Optimization:** Given the rule $X_t(\mu_t^*) = \mathcal{E}_t / p_t(\mu_t^*)$, the household must choose expenditures per period. He solves

$$\begin{aligned} \max_{\{E_t\}} \quad & \sum_{t=0}^{\infty} \beta^t \log \left(\lambda^{\mu_t^*} \mathcal{E}_t \frac{1}{p_t(\mu_t^*)} \right) \\ \text{s.t.} \quad & \sum_{t=0}^{\infty} \beta^t \mathcal{E}_t \leq \frac{1}{1-\beta} \bar{Y}. \end{aligned}$$

Rewriting

$$\log \left(\lambda^{\mu_t^*} \mathcal{E}_t \frac{1}{p_t(\mu_t^*)} \right) = \mu_t^* \log \lambda + \log \mathcal{E}_t - \log p_t(\mu_t^*),$$

we get the FONC $1/\mathcal{E}_t = v$, where v is the Lagrange multiplier on the budget constraint. It then follows that E_t is constant across all t , and hence $\mathcal{E}_t = \bar{Y}$.

²We are assuming in the budget constraint that the gross one-period rate of return $R = \beta^{-1}$.

We end with the very simple demand functions

$$\begin{aligned} X_t^*(\mu_t) &= \begin{cases} \bar{Y}/p(\mu_t), & p(\mu_t) \leq \lambda^{\mu_t - \mu_{t-1}} p(\mu_{t-1}) \\ 0, & p(\mu_t) > \lambda^{\mu_t - \mu_{t-1}} p(\mu_{t-1}) \end{cases} \\ X_t^*(\mu_{t-1}) &= \begin{cases} 0, & p(\mu_t) \leq \lambda^{\mu_t - \mu_{t-1}} p(\mu_{t-1}) \\ \bar{Y}/p(\mu_{t-1}), & p(\mu_t) > \lambda^{\mu_t - \mu_{t-1}} p(\mu_{t-1}) \end{cases} \\ X_t^*(\mu) &= 0, \quad \mu < \mu_{t-1} \end{aligned}$$

B.1.1.2 Utility Specification #2 (Money in the Utility Function)

The standard utility function in much of the partial-equilibrium patent literature (see O'Donoghue et al. [1998], Hopenhayn and Mitchell [2010], O'Donoghue [1998], Rockett [2010]) would be

$$U \equiv \mathbf{E}_0 \sum_{t=0}^{\infty} \beta^t u \left(\sum_{\mu=0}^{\mu_t} [\lambda(\mu+1) - p_t(\mu)] X_t(\mu) \right), \quad (\text{B.1})$$

where in each period we restrict the household to unit demand, that is $\sum_{\mu=0}^{\mu_t} X_t(\mu) = 1$. It is clear that if we define $\mu_t^* = \arg \max_{\mu \leq \mu_t} \{\lambda(\mu+1) - p_t(\mu)\}$ the consumer has demand function $X_t^*(\mu) = \mathbf{1}_{\{\mu = \mu_t^*\}}$.

B.1.2 Pricing and Indirect Utility

All firms price qualities $\mu < \mu_t$ at marginal cost, c .

B.1.2.1 Utility Specification #1

The monopolist captures the entire market so long as

$$\frac{\lambda^{\mu_{t-1} + 1}}{p_t(\mu_t)} \geq \frac{\lambda^{\mu_{t-1}}}{c},$$

that is $p_t(\mu_t) \leq \lambda^l c$. The monopolist chooses $p_t(\mu_t) = \lambda^l c$.³ Under this specification, the monopolist receives market profits

$$\pi = \frac{\bar{Y}}{\lambda^l c} (\lambda^l c - c) = \bar{Y} \left(1 - \frac{1}{\lambda^l} \right),$$

and non-patent-holders get zero. We use $\pi(\chi, \tau, l)$ to denote profits to the firm. This yields the indirect utility index

$$\begin{aligned} u_t &= \log \left(\lambda^{\mu_t} \frac{\bar{Y}}{p_t(\mu_t)} \right) = \mu_t \log \lambda + \log \bar{Y} - l \log \lambda - \log c \\ &= \mu_t \log \lambda + \log \bar{Y} - (\mu_t - \mu_{t-1}) \log \lambda - \log c \\ &= \mu_{t-1} \log \lambda + \log \bar{Y} - \log c \end{aligned}$$

and so the indirect utility function is

$$U = \log \lambda \sum_{t=0}^{\infty} \beta^t \mathbf{E}[\mu_{t-1}] + \frac{\log(\bar{Y}/c)}{1-\beta}.$$

Note that U is finite as

$$\sum_{t=0}^{\infty} \beta^t \mathbf{E}[\mu_{t-1}] \leq \sum_{t=0}^{\infty} \beta^t [\Delta t] = \Delta \sum_{t=1}^{\infty} \beta^t \sum_{s=0}^{\infty} \beta^s = \frac{\Delta \beta}{(1-\beta)^2}.$$

Utility Specification #2 The monopolist captures the entire market so long as

$$\lambda (\mu_{t-1} + l) \mu_t - p_t(\mu_t) \geq \lambda \mu_{t-1} - c,$$

that is

$$p_t(\mu_t) \geq \lambda l + c$$

³There is some indeterminacy in that the consumer is indifferent between qualities μ_t and μ_{t-1} when $p_t(\mu_t) = \lambda^l p_t(\mu_t)$. So in theory, we can assume that $p_t(\mu_t) = \lambda^l c - \varepsilon$ for infinitesimal ε to break the tie.

The monopolist sets price $p_t(\mu_t) = \lambda \iota + c$ and gets market profits $\pi = \lambda \iota$. Non-patent-holders get zero. The indirect utility function is

$$\begin{aligned} U &= \sum_{t=0}^{\infty} \beta^t u(\lambda (\mu_{t-1} + \iota) \mu_t - (\lambda \iota + c)) \\ &= \sum_{t=0}^{\infty} \beta^t u(\lambda (\mu_t^2 - \mu_t + \mu_{t-1}) - c). \end{aligned}$$

U will be bounded for a good choice of u .

B.2 The State Transition Probabilities in 3

We will view this as if the state space is from the perspective of firm #1. Let (χ, τ, ι) be the current state and let I^{-1} denote the actions of firm one's competitors. We note that the state is implicit in the function Q and drop it for simplicity. The transitions as a function of the firms choice I_1 are as follows. Recall that

$$Q_{\delta}(I_i | \dots) = \begin{cases} 1 - \phi(I_i | \dots), & \delta = 0 \\ \phi(I_i | \dots) [1 - \eta(I_i | \dots)], & \delta = 1 \\ \phi(I_i | \dots) \eta(I_i | \dots), & \delta = \Delta \end{cases}.$$

If $\iota = \Delta$ and $\tau > 1$ The transitions when only a breakthrough patent holder is allowed to innovate are quite simple. Let j be the identity of the patent holder (possibly $j = 1$). Then

$$(\chi, \tau, \Delta) \rightarrow \begin{cases} (\chi, [\tau - 1]^+, \Delta), & \text{w.p. } Q_0(I_j) + \Xi_{\tau} Q_1(I_j) \\ (\chi, T, 1), & \text{w.p. } (1 - \Xi_{\tau}) Q_1(I_j) \\ (\chi, T, \Delta), & \text{w.p. } Q_{\Delta}(I_j) \end{cases},$$

where $\Xi_{\tau} \in \{0, 1\}$ denotes whether incremental innovations are registered by a patent holder. These are dependent on the value function, and varying in τ .

If $\iota = 1$ or ($\iota = \Delta$ and $\tau \leq 1$) This case is considerably more cumbersome, as we can move to one of five states.

The probability of no innovation is obviously

$$P(\chi' = \chi, \tau' = [\tau - 1]^+, \iota') = Q_0(I_1) \prod_{j=2}^N Q_0(I_j).$$

The probability that at least one other firm has a breakthrough is $1 - \prod_{j=2}^N (1 - Q_\Delta(I_j))$ so the probability that another firm is awarded a breakthrough patent conditional on firm 1 not having a breakthrough is $1 - \prod_{j=2}^N (1 - Q_\Delta(I_j))$. The probability that another firm is awarded a breakthrough patent conditional on firm 1 having a breakthrough is

$$\sum_{k=0}^{N-1} \frac{k}{1+k} \sum_{\substack{\mathcal{A} \subseteq N \setminus \{1\} \\ \#\mathcal{A}=k}} \prod_{j \in \mathcal{A}} Q_\Delta(I_j) \prod_{\ell \notin \mathcal{A} \setminus \{1\}} (1 - Q_\Delta(I_\ell)).$$

So the probability that the state goes to $(1, T, \Delta)$ is

$$\begin{aligned} P(1, T, \Delta) &= (1 - Q_\Delta(I_1)) \left[1 - \prod_{j=2}^N (1 - Q_\Delta(I_j)) \right] \\ &\quad + Q_\Delta(I_1) \sum_{k=0}^{N-1} \frac{k}{1+k} \sum_{\substack{\mathcal{A} \subseteq N \setminus \{1\} \\ \#\mathcal{A}=k}} \prod_{j \in \mathcal{A}} Q_\Delta(I_j) \prod_{\ell \notin \mathcal{A} \setminus \{1\}} (1 - Q_\Delta(I_\ell)). \end{aligned}$$

So the probability that firm 1 is awarded a breakthrough patent is

$$P(0, T, \Delta) = Q_\Delta(I_1) \sum_{k=0}^{N-1} \frac{1}{1+k} \sum_{\substack{\mathcal{A} \subseteq N \setminus \{1\} \\ \#\mathcal{A}=k}} \prod_{j \in \mathcal{A}} Q_\Delta(I_j) \prod_{\ell \notin \mathcal{A} \setminus \{1\}} (1 - Q_\Delta(I_\ell)).$$

Recall that an incremental patent is awarded if there is an incremental innovation *but no breakthrough*. The probability of no breakthrough is $(1 - Q_\Delta(I_1)) \prod_{j=2}^N (1 - Q_\Delta(I_j))$.

The probability that firm 1 is awarded an incremental patent is

$$\begin{aligned}
P(0, T, 1) &= \prod_{j=2}^N (1 - Q_\Delta(I_j)) \cdot \\
&\quad Q_1(I_1) \sum_{k=0}^{N-1} \frac{1}{1+k} \sum_{\substack{\mathcal{A} \subseteq N \setminus \{1\} \\ \#\mathcal{A}=k}} \prod_{j \in \mathcal{A}} \frac{Q_1(I_j)}{1 - Q_\Delta(I_j)} \prod_{\ell \notin \mathcal{A} \cup \{1\}} \frac{Q_0(I_\ell)}{1 - Q_\Delta(I_\ell)} \\
&= Q_1(I_1) \sum_{k=0}^{N-1} \frac{1}{1+k} \sum_{\substack{\mathcal{A} \subseteq N \setminus \{1\} \\ \#\mathcal{A}=k}} \prod_{j \in \mathcal{A}} Q_1(I_j) \prod_{\ell \notin \mathcal{A} \cup \{1\}} Q_0(I_\ell)
\end{aligned}$$

Note that we have conditioned on the fact that no firm $j = 2, \dots, N$ has had a breakthrough. This is why we have the term $\prod_{\ell \notin \mathcal{A} \cup \{1\}} Q_0(I_\ell)$ and not $\prod_{\ell \notin \mathcal{A} \cup \{1\}} (1 - Q_1(I_\ell))$.

Condition on firm 1 being unsuccessful (w.p. $Q_0(I_1)$); the other $N - 1$ firms do *not* get an incremental patent between them if (i) one of them has a breakthrough or (ii) none of them have any innovation. (i) and (ii) are disjoint. The probability of (i) is $1 - \prod_{j=2}^N (1 - Q_\Delta(I_j))$. The probability of (ii) is $\prod_{j=2}^N Q_0(I_j)$. So the probability that another firm gets an incremental patent (conditional on firm 1 being unsuccessful) is

$$\begin{aligned}
&Q_0(I_1) \left[1 - \left(\left[1 - \prod_{j=2}^N (1 - Q_\Delta(I_j)) \right] + \prod_{j=2}^N Q_0(I_j) \right) \right] \\
&= Q_0(I_1) \left[\prod_{j=2}^N (1 - Q_\Delta(I_j)) - \prod_{j=2}^N Q_0(I_j) \right]
\end{aligned}$$

Therefore, the probability that some firm that isn't 1 is awarded an incremental patent is simply

$$\begin{aligned}
P(1, T, 1) &= Q_0(I_1) \left[\prod_{j=2}^N (1 - Q_\Delta(I_j)) - \prod_{j=2}^N Q_0(I_j) \right] \\
&\quad + Q_1(I_1) \sum_{k=0}^{N-1} \frac{k}{1+k} \sum_{\substack{\mathcal{A} \subseteq N \setminus \{1\} \\ \#\mathcal{A}=k}} \prod_{j \in \mathcal{A}} Q_1(I_j) \prod_{\ell \notin \mathcal{A} \cup \{1\}} Q_0(I_\ell).
\end{aligned}$$

B.2.1 Transition Probabilities Facing a Firm in a Symmetric MPE

Now, assume that the $N - 1$ competitors of firm 1 use a symmetric strategy denoted $I_{\chi, \tau, \iota}^*$. Let Ξ_τ^* denote the incremental innovation registration decision used by competitor breakthrough patent-holders. This simplifies the above expressions tremendously. We use the following expressions in our computations.

When $\iota = \Delta$ and $\tau \geq 2$ The transitions when $\iota = \Delta$ and $\tau \geq 2$ can be left basically unchanged. We write when firm 1 has the patent ($\chi = 0$),

$$(0, \tau, \Delta) \rightarrow \begin{cases} (0, [\tau - 1]^+, \Delta), & \text{w.p. } Q_0(I_1) + \Xi_\tau Q_1(I_1) \\ (0, T, 1), & \text{w.p. } (1 - \Xi_\tau) Q_1(I_1) \\ (0, T, \Delta), & \text{w.p. } Q_\Delta(I_1) \end{cases},$$

and when firm 1 doesn't have the patent ($\chi = 1$),

$$(1, \tau, \Delta) \rightarrow \begin{cases} (0, [\tau - 1]^+, \Delta), & \text{w.p. } Q_0(I_{0, \tau, \Delta}^*) + \Xi_\tau^* Q_1(I_{0, \tau, \Delta}^*) \\ (0, T, 1), & \text{w.p. } (1 - \Xi_\tau^*) Q_1(I_{0, \tau, \Delta}^*) \\ (0, T, \Delta), & \text{w.p. } Q_\Delta(I_{0, \tau, \Delta}^*) \end{cases}.$$

When $\iota = 1$ or ($\iota = \Delta$ and $\tau \leq 2$) and $\chi = 0$:

$$P(0, [\tau - 1]^+, \iota) = Q_0(I_1) [Q_0(I_{1, \tau, \iota}^*)]^{N-1}$$

$$\begin{aligned} P(1, T, \Delta) &= (1 - Q_\Delta(I_1)) \left[1 - (1 - Q_\Delta(I_{1, \tau, \iota}^*))^{N-1} \right] \\ &\quad + Q_\Delta(I_1) \sum_{k=0}^{N-1} \frac{k}{1+k} \binom{N-1}{k} [Q_\Delta(I_{1, \tau, \iota}^*)]^k [1 - Q_\Delta(I_{1, \tau, \iota}^*)]^{N-1-k} \end{aligned}$$

$$P(0, T, \Delta) = \mathcal{Q}_\Delta(I_1) \sum_{k=0}^{N-1} \frac{1}{1+k} \binom{N-1}{k} [\mathcal{Q}_\Delta(I_{1,\tau,\iota}^*)]^k [1 - \mathcal{Q}_\Delta(I_{1,\tau,\iota}^*)]^{N-1-k}$$

$$\begin{aligned} P(1, T, 1) &= \mathcal{Q}_0(I_1) \left[[1 - \mathcal{Q}_\Delta(I_{1,\tau,\iota}^*)]^{N-1} - [\mathcal{Q}_0(I_{1,\tau,\iota}^*)]^{N-1} \right] \\ &\quad + \mathcal{Q}_1(I_1) \sum_{k=0}^{N-1} \frac{k}{1+k} \binom{N-1}{k} [\mathcal{Q}_1(I_{1,\tau,\iota}^*)]^k [\mathcal{Q}_0(I_{1,\tau,\iota}^*)]^{N-1-k} \end{aligned}$$

$$P(0, T, 1) = \mathcal{Q}_1(I_1) \sum_{k=0}^{N-1} \frac{1}{1+k} \binom{N-1}{k} [\mathcal{Q}_1(I_{1,\tau,\iota}^*)]^k [\mathcal{Q}_0(I_{1,\tau,\iota}^*)]^{N-1-k}$$

When $\iota = 1$ or ($\iota = \Delta$ and $\tau \leq 2$) and $\chi = 1$:

$$P(1, [\tau - 1]^+, \iota) = \mathcal{Q}_0(I_1) \mathcal{Q}_0(I_{0,\tau,\iota}^*) [\mathcal{Q}_0(I_{1,\tau,\iota}^*)]^{N-2}$$

$$\begin{aligned} P(1, T, \Delta) &= (1 - \mathcal{Q}_\Delta(I_1)) \left[1 - (1 - \mathcal{Q}_\Delta(I_{0,\tau,\iota}^*)) (1 - \mathcal{Q}_\Delta(I_{1,\tau,\iota}^*))^{N-2} \right] \\ &\quad + \mathcal{Q}_\Delta(I_1) \sum_{k=0}^{N-2} \left\{ \left(\frac{\mathcal{Q}_\Delta(I_{0,\tau,\iota}^*) (k+1)}{k+2} + \frac{(1 - \mathcal{Q}_\Delta(I_{0,\tau,\iota}^*)) k}{k+1} \right) \right. \\ &\quad \left. \times \binom{N-2}{k} [\mathcal{Q}_\Delta(I_{1,\tau,\iota}^*)]^k [1 - \mathcal{Q}_\Delta(I_{1,\tau,\iota}^*)]^{N-2-k} \right\}. \end{aligned}$$

$$\begin{aligned} P(0, T, \Delta) &= \mathcal{Q}_\Delta(I_1) \sum_{k=0}^{N-2} \left(\frac{\mathcal{Q}_\Delta(I_{0,\tau,\iota}^*)}{k+2} + \frac{1 - \mathcal{Q}_\Delta(I_{0,\tau,\iota}^*)}{k+1} \right) \\ &\quad \times \binom{N-2}{k} [\mathcal{Q}_\Delta(I_{1,\tau,\iota}^*)]^k [1 - \mathcal{Q}_\Delta(I_{1,\tau,\iota}^*)]^{N-2-k} \end{aligned}$$

$$\begin{aligned}
P(1, T, 1) &= Q_0(I_1) \left[(1 - Q_\Delta(I_{0,\tau,\iota}^*)) (1 - Q_\Delta(I_{1,\tau,\iota}^*))^{N-2} - Q_0(I_{0,\tau,\iota}^*) (Q_0(I_{1,\tau,\iota}^*))^{N-2} \right] \\
&\quad + Q_1(I_1) \sum_{k=0}^{N-2} \left(\frac{Q_0(I_{0,\tau,\iota}^*) k}{1+k} + \frac{Q_1(I_{0,\tau,\iota}^*) (1+k)}{2+k} \right) \\
&\quad \times \binom{N-2}{k} [Q_1(I_{1,\tau,\iota}^*)]^k [Q_0(I_{1,\tau,\iota}^*)]^{N-2-k}
\end{aligned}$$

$$\begin{aligned}
P(0, T, 1) &= Q_1(I_1) \sum_{k=0}^{N-1} \left(\frac{Q_0(I_{0,\tau,\iota}^*)}{1+k} + \frac{Q_1(I_{0,\tau,\iota}^*)}{2+k} \right) \\
&\quad \times \binom{N-2}{k} [Q_1(I_{1,\tau,\iota}^*)]^k [Q_0(I_{1,\tau,\iota}^*)]^{N-2-k}
\end{aligned}$$

B.3 Optimal Pure Strategy Investment in 3

We consider the special case that was simulated in the numerical section of the paper, where

$$\phi(I) = aI, \quad \eta(I) = \bar{\eta}$$

where $a \in (0, 1]$ and $\bar{\eta} \in (0, 1)$. We again take C as strictly increasing and strictly convex with $C(0) = 0$. The intensity-invariant breakthrough probability $\bar{\eta}$ and linear success probability ϕ . (The case of a general (C, ϕ, η) is simply too complicated to discuss.)

Claim B.1. For any value function V , competitor intensity I^{-i} , and state (χ, τ, ι) , the firm has a unique optimal intensity $I^* \in [0, 1]$ (a unique best response).

Proof. Note that under this functional specification,

$$Q_\delta(I) = \begin{cases} 1 - aI, & \delta = 0 \\ aI(1 - \bar{\eta}), & \delta = 1 \\ aI\bar{\eta}, & \delta = \Delta \end{cases}$$

Note that the function

$$R(I) := \mathbf{E} [V(\chi', \tau', \iota') \mid I, I^{-i}, \chi, \tau, \iota]$$

is linear in I for any value function V , competitor intensity I^{-i} and state (χ, τ, ω) . That is

$$R(I) = b_0(I^{-i}, V, \chi, \tau, \iota) + b_1(I^{-i}, V, \chi, \tau, \iota)I.$$

This follows because

$$R(I) = \sum_{\chi'} \sum_{\tau'} \sum_{\iota'} V(\chi', \tau', \iota') \Pr\{\chi', \tau', \iota' \mid I, I^{-i}, \chi, \tau, \iota\}$$

is and $\Pr\{\chi', \tau', \iota' \mid I, I^{-i}, \chi, \tau, \iota\}$ is linear in I for all (χ', τ', ι') .⁴ Fixing the actions of competitors and the state, the marginal benefit of intensity (at any intensity) is just some weighted sum of the value function across states.

So we can write the marginal cost of investment as $C'(I)$ and the marginal benefit as $R'(I) = b_1(I^{-i}, \chi, \tau, \iota)$. We then get the following unique investment decisions:

$$I^*(\chi, \tau, \iota \mid I^{-i}, V) = \begin{cases} 0, & b_1(I^{-i}, \chi, \tau, \iota) \leq C'(0) \\ (C')^{-1}(b_1(I^{-i}, \chi, \tau, \iota)), & b_1(I^{-i}, \chi, \tau, \iota) \in (C'(0), C'(1)) \\ 1, & b_1(I^{-i}, \chi, \tau, \iota) \geq C'(1) \end{cases}$$

⁴In fact for most states (χ', τ', ι') , $\Pr\{\chi', \tau', \iota' \mid I, I^{-i}, \chi, \tau, \iota\} = 0$. As an example of when this is *not* the case, note that we can write

$$\Pr\{0, T, 1 \mid I, I^{-i}, \chi, \tau, \iota = 1\} = aI(1 - \bar{\eta}) \sum_{k=0}^{N-1} \frac{1}{1+k} \sum_{\substack{\mathcal{A} \subseteq N \setminus \{1\} \\ \#\mathcal{A}=k}} [a(1 - \bar{\eta})]^k \prod_{j \in \mathcal{A}} I_j^{-i} \prod_{\ell \notin \mathcal{A} \cup \{1\}} (1 - aI_\ell^{-i})$$

so in this case,

$$b_1(I^{-i}, V, \chi, \tau, \iota) = \Gamma(I^{-i}, V, \chi, \tau, \iota) + a(1 - \bar{\eta}) \sum_{k=0}^{N-1} \frac{1}{1+k} \sum_{\substack{\mathcal{A} \subseteq N \setminus \{1\} \\ \#\mathcal{A}=k}} [a(1 - \bar{\eta})]^k \prod_{j \in \mathcal{A}} I_j^{-i} \prod_{\ell \notin \mathcal{A} \cup \{1\}} (1 - aI_\ell^{-i}) \cdot V(0, T, 1),$$

where $\Gamma(I^{-i}, V, \chi, \tau, \iota)$ is some constant.

Since C is strictly convex, $(C')^{-1}$ is a well-defined injective function.

□

Note that if we take C' to have range $[0, \infty)$ (that is the marginal cost of intensity becomes unbounded as innovation becomes a certainty) we get the simpler rule

$$I^*(\chi, \tau, \iota | I^{-i}, V) = \begin{cases} 0, & b_1(I^{-i}, \chi, \tau, \iota) < 0 \\ (C')^{-1}(b_1(I^{-i}, \chi, \tau, \iota)), & b_1(I^{-i}, \chi, \tau, \iota) \geq 0 \end{cases}.$$

Moreover, if we assume V such that $b_1(I^{-i}, V, \chi, \tau, \iota) \geq 0$ for all I^{-i} and states (marginal benefit of investing is always non-negative), we get the rule

$$I^*(\chi, \tau, \iota | I^{-i}, V) = (C')^{-1}(b_1(I^{-i}, \chi, \tau, \iota)).$$

The MPE value function V *should* give this to us, but the argument in section B.4 will require a unique intensity choice for *any* value function. So could easily input bizarre value functions $\tilde{V}$ that would give you negative returns to intensity.

B.4 Existence of a Symmetric Markov-Perfect Equilibrium in

3

We follow Doraszelski and Satterthwaite [2007] in establishing the existence of a symmetric MPE for our simpler model. Consider the following assumptions:

Assumption B.2. (i) *The state space is finite.* (ii) *Profits are bounded (i.e. there exists $\underline{\Pi}, \bar{\Pi}$ such that $-\infty < \underline{\Pi} < \pi(\chi, \tau, \iota) < \bar{\Pi} < \infty$).* (iii) *Investments are bounded.* (iv) *Firms discount future payoffs, i.e. $\beta \in [0, 1)$.*

Our model satisfies all of these assumptions, as profits are contained in a finite set that is stationary and firms discount according to β . Assumption (ii) requires some mention, as C can

be unbounded as $I \rightarrow 1$ (investment is $C(I)$, not I). For C unbounded⁵ and for any (N, T) we can always bound investment by

$$I \leq I_{\max} = C^{-1} \left(\frac{\beta}{1-\beta} \pi(0, T, \Delta) \right),$$

this is the firm will never invest more today than the net present value of being a breakthrough monopolist forever. In fact, this is a very loose bound; firms will invest significantly less.

Let

$$h_i(\chi, \tau, \iota, I_i, I_{-i}V) = \pi(\chi, \tau, \iota) - C(I) + \beta \mathbf{E} [V_i(\chi', \tau', \iota') \mid I_i, I_{-i}, \chi, \tau, \iota],$$

so that

$$V_i(\chi, \tau, \iota) = \max_{I_i} h_i(\chi, \tau, \iota, I_i, I_{-i}V_i).$$

h_i is called the *local income function*.

Assumption B.3. h_i is a continuous function in I_i for all states (χ, τ, ι) and value functions V_i .

This assumption is satisfied for any value function V_i and profit function $\pi(\cdot)$. V_i enters only through the expectation, and π is additively separable and does not include I_i . We only require that the transition probabilities be continuous in I_i (which they are given the continuity of η, ϕ).

Assumption B.4. There is a unique $I^*(\chi, \tau, \iota)$ that maximizes $h_i(\chi, \tau, \iota, I_i, I_{-i}V_i)$ for any state, competitor action and value function V_i .

See appendix B.3 for a proof of this for the specification used in the simulations.

A function $f : A^N \rightarrow B^N$ is *symmetric* if for any permutation σ on $\{1, \dots, N\}$,

$$f_n(x_1, \dots, x_N) = f_{\sigma(n)}(x_{\sigma(1)}, \dots, x_{\sigma(N)}).$$

Assumption B.5. The local income function $(h_1, \dots, h_N)$ are symmetric.

⁵If C is bounded we have no issue.

	T^*	$N(T^*)$	$\mathbf{E}[\mu' - \mu \mid T^*]$	$\mathbf{E}[U \mid T^*]$
$N = 2$	8	-	.073	60.356
$N = 7$	4	-	.198	64.503
Endogenous $N(T)$	12	10	.223	65.100

Table B.1: Duopoly $N = 2$ vs. Endogenous $N(T)$

This is obvious given the lack of firm-dependence on any of the primitive functions we have written out; payoffs and transitions vary only in states. It is helpful to think of the state as

$$\omega = (\omega_1, \dots, \omega_N) = ((\chi_1, \tau, \iota), \dots, (\chi_N, \tau, \iota)),$$

where χ_i denotes whether i is a patent-holder.

Proposition. *Under assumptions B.2, B.3, B.4, B.5, a symmetric Markov-perfect equilibrium exists in pure investment strategies.*

Proof. See Doraszelski and Satterthwaite [2007], Propositions 1, 3 and 5. □

B.5 Free Entry Matters in 3

We provide a brief example to show how radically solutions differ between exogenous and endogenous N . We use the standard baseline parameter values from the main paper; $\Delta = 3$, $\lambda = 1.1$, $\beta = .95$, $\bar{Y} = 5$, $x^e = .9$, $c = .3$, $\phi(I) = .35I$, $\eta(I) = .2$ and $C(I) = 500I^4/(1 - I)$. We first fix N at 2 (duopoly) and then consider $N(T)$.

We see that the policy implications differ depending on whether the market concentration is fixed. When it is not fixed, we would intuitively want longer patent lengths to encourage entrants and competition.

B.6 Using the optimalpatents package from 3

The optimalpatents Java package can be downloaded at

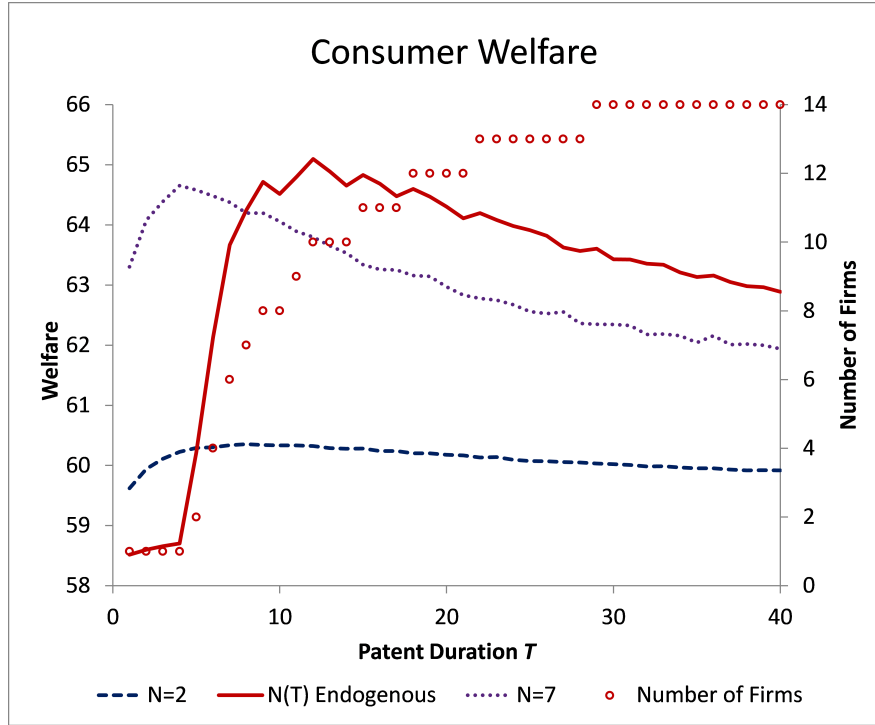


Figure B.1: Duopoly $N = 2$ vs. Endogenous $N(T)$

<http://www.princeton.edu/~rkatzwer/optimalpatents/optimalpatentscode.zip>.

The zip file includes a copy of the Apache Common's Math library which is used by the package.

This is a guide to using the `optimalpatents` package. The package contains three main classes; `Specification` (which defines the primitive parameters and functional forms of the model), `ValueFunction` (a data structure which stores the value- and policy-functions), and `ValueFunctionSolver` (which contains the machinery of the symmetric MPE solution iteration). Below, all variables are exactly as in the paper, but note that in code $\beta \equiv \text{beta}$, etc.

Specification A *specification* of the model can be customized by extending the `Specification` class. The constructor is

```
Specification(double  $\beta$ , double  $\lambda$ , double  $c$ , double  $\Delta$ ,
double  $x^f$ , int  $N$ ) {
    ...
}
```

```
}
```

Where these parameters are exactly as in the paper above. One can call the `setNumberFirms(int  $N$ )` method to set the number of firms.

One must define functional forms for the following abstract method signatures:

```
double eta(double  $I_i$ , double  $\sum_{j \neq i} I_j$ , int  $\chi$ , int  $\tau$ , int  $\iota$ );
double phi(double  $I_i$ , double  $\sum_{j \neq i} I_j$ , int  $\chi$ , int  $\tau$ , int  $\iota$ );
double costInnov(double  $I_i$ , double  $\sum_{j \neq i} I_j$ , int  $\chi$ , int  $\tau$ , int  $\iota$ );
double pi(int  $\chi$ , int  $\tau$ , int  $\iota$ );
double quantity(double  $p(\mu_t)$ );
double price(int  $\tau$ , int  $\iota$ );
double utilityIndex(double  $X(\mu_t)$ , double  $\mu_t$ , double  $p(\mu_t)$ );
```

Respectively, these are the η , ϕ , C , π , $X(\mu_t)$, $p(\mu_t)$ and u functions from the paper. (The specification used in section 3.3 can be instantiated via the `PaperSpecification` class in the `optimalpatents` package. There is a static method `PaperSpecification.getBaseline()` that returns the baseline parametrization.)

Value Function The value function class stores a value function and policy function. The constructor

```
new ValueFunction(int  $N$ , int  $T$ ) { ... }
```

instantiates an empty value function with patent duration T . The methods

```
void setPolicy(int  $\chi$ , int  $\tau$ , int  $\iota$ , double  $i$ ) { ... }
void setValue(int  $\chi$ , int  $\tau$ , int  $\iota$ , double  $v$ ) { ... }
```

set $V(\chi, \tau, \iota) = v$ and $I^*(\chi, \tau, \iota) = i$, respectively. The values stores in the value function can then be called using

```
double policy(int  $\chi$ , int  $\tau$ , int  $\iota$ ) { ... }
double value(int  $\chi$ , int  $\tau$ , int  $\iota$ ) { ... }
```

The method

```
public SimulationResult simulate(int numPeriods, int
numDraws, Specification specs) {
    ...
}
```

returns a data structure, `SimulationResult`, that contains the results of the model simulated under the value function with model specification `specs`. The method will simulate out to `numPeriods` many periods, with `numDraws` many samples. `SimulationResult` contains statistics such as expected discounted utility, average R&D spending, $\mathbf{E}[\mu_{t+1} - \mu_t]$, etc.

`ValueFunctionSolver` This class contains the machinery for solving for MPEs. It has two major static methods:

```
ValueFunction solveFixedN(Specification specs, int  $N$ , int  $T$ ) { ... }
ValueFunction solveEndogenousN(Specification specs,
int  $N_{\text{initial}}$ , int  $T$ ) { ... }
```

The first method solve for the symmetric MPE value function for a fixed number of firms N and patent policy T . The second method solves for the number of firms $N(T)$ that approximates the zero-economic-profit condition (N_{initial} is the user supplied guess of $N(T)$) and returns the MPE value function.

The class also contains the more flexible methods that allow the user to specify the maximum allowable iterations, a convergence criterion, Gauss-Jacobi vs. Gauss-Seidel updating, etc.

B.7 Comparative Statics Tables in 3

Low Entry Cost ($x^f = .0225$)								
Δ	T^*	$N(T^*)$	$\mathbf{E}[\mu_{t+1} - \mu_t]$	$\mathcal{T}_{\min}$	$N(\mathcal{T}_{\min})$	R&D	FC	Π
1	10	11	.153	4	3	1.50	4.95	.15
1.5	10	14	.180	4	5	1.97	6.30	.19
2	7	15	.257	4	7	2.15	6.75	.35
2.5	7	18	.289	3	3	2.61	8.10	.42
3	6	20	.383	3	5	2.93	9.00	.30
3.5	6	23	.416	3	8	3.43	10.35	.51
4	6	27	.527	3	10	4.02	12.15	.05

Medium Entry Cost ($x^f = .045$)								
Δ	T^*	$N(T^*)$	$\mathbf{E}[\mu_{t+1} - \mu_t]$	$\mathcal{T}_{\min}$	$N(\mathcal{T}_{\min})$	R&D	FC	Π
1	17	5	.088	7	1	1.25	4.50	.15
1.5	15	6	.103	5	1	1.55	5.40	.22
2	17	8	.145	6	2	2.12	7.20	.17
2.5	14	9	.167	5	1	2.45	8.10	.33
3	12	10	.223	6	4	2.78	9.00	.30
3.5	12	10	.223	5	2	2.77	9.00	.37
4	11	13	.309	5	4	3.72	11.70	.37

High Entry Cost ($x^f = .0675$)								
Δ	T^*	$N(T^*)$	$\mathbf{E}[\mu_{t+1} - \mu_t]$	$\mathcal{T}_{\min}$	$N(\mathcal{T}_{\min})$	R&D	FC	Π
1	23	3	.059	11	1	.97	4.05	.11
1.5	26	4	.070	10	1	1.39	5.40	.13
2	25	5	.100	9	1	1.85	6.75	.09
2.5	23	6	.115	8	1	2.30	8.10	.15
3	21	7	.160	7	1	2.75	9.45	.24
3.5	16	7	.173	7	2	2.86	9.45	.41
4	15	8	.221	7	2	3.31	10.80	.26

Table B.2: Optimal policy as a function of breakthrough size Δ

Low Entry Cost ($x^f = .0225$)								
$\bar{\eta}$	T^*	$N(T^*)$	$\mathbf{E}[\mu_{t+1} - \mu_t]$	$\mathcal{T}_{\min}$	$N(\mathcal{T}_{\min})$	R&D	FC	Π
.10	11	20	.317	3	1	2.83	9.00	.11
.15	7	19	.350	3	3	2.75	8.55	.30
.20	6	20	.383	3	5	2.92	9.00	.36
.25	6	22	.410	3	8	3.29	9.90	.46
.30	5	21	.438	3	10	3.19	9.45	.67
.35	5	23	.465	3	12	3.51	10.35	.65
.40	4	20	.485	2	1	3.15	9.00	.65

Medium Entry Cost ($x^f = .045$)								
$\bar{\eta}$	T^*	$N(T^*)$	$\mathbf{E}[\mu_{t+1} - \mu_t]$	$\mathcal{T}_{\min}$	$N(\mathcal{T}_{\min})$	R&D	FC	Π
.10	20	9	.180	6	2	2.37	8.10	.19
.15	16	10	.202	6	3	2.74	9.00	.10
.20	12	10	.223	6	4	2.77	9.00	.41
.25	10	10	.240	6	5	2.86	9.00	.33
.30	10	11	.255	4	1	3.19	9.90	.46
.35	8	10	.298	4	1	2.98	9.00	.67
.40	7	10	.296	4	2	2.96	9.00	.53

High Entry Cost ($x^f = .0675$)								
$\bar{\eta}$	T^*	$N(T^*)$	$\mathbf{E}[\mu_{t+1} - \mu_t]$	$\mathcal{T}_{\min}$	$N(\mathcal{T}_{\min})$	R&D	FC	Π
.10	48	6	.110	9	1	2.25	8.10	.02
.15	22	6	.140	8	1	2.31	8.10	.21
.20	21	7	.167	7	1	2.76	9.45	.22
.25	13	6	.170	7	2	2.43	8.10	.18
.30	14	7	.183	6	1	2.91	9.45	.33
.35	12	7	.191	6	2	2.95	9.45	.34
.40	11	7	.209	6	2	3.01	9.45	.43

Table B.3: Optimal policy as a function of breakthrough probability $\bar{\eta}$

Low Entry Cost ($x^f = .0225$)								
λ	T^*	$N(T^*)$	$\mathbf{E}[\mu_{t+1} - \mu_t]$	$\mathcal{T}_{\min}$	$N(\mathcal{T}_{\min})$	R&D	FC	Π
1.05	14	11	.201	6	3	1.53	4.95	.09
1.10	6	20	.383	3	5	2.93	9.00	.29
1.15	4	27	.520	2	6	4.08	12.15	.75
1.20	3	33	.637	2	19	5.03	14.85	.92
1.25	3	44	.634	2	31	8.43	19.80	.99

Medium Entry Cost ($x^f = .045$)								
λ	T^*	$N(T^*)$	$\mathbf{E}[\mu_{t+1} - \mu_t]$	$\mathcal{T}_{\min}$	$N(\mathcal{T}_{\min})$	R&D	FC	Π
1.05	31	5	.102	7	1	1.27	4.50	.05
1.10	12	10	.223	4	1	1.15	9.00	.31
1.15	7	13	.291	3	1	3.22	9.90	.80
1.20	6	17	.387	2	1	5.07	15.30	1.09
1.25	6	23	.453	3	12	6.75	20.70	.98

High Entry Cost ($x^f = .0675$)								
λ	T^*	$N(T^*)$	$\mathbf{E}[\mu_{t+1} - \mu_t]$	$\mathcal{T}_{\min}$	$N(\mathcal{T}_{\min})$	R&D	FC	Π
1.05	27	2	.059	19	1	1.37	2.70	.05
1.10	21	7	.154	7	1	2.76	9.45	.20
1.15	13	10	.229	5	2	4.94	13.50	.35
1.20	9	12	.291	4	3	5.05	16.20	.67
1.25	7	13	.334	3	2	5.74	17.55	1.28

Table B.4: Optimal policy as a function of quality step λ

Low Entry Cost ($x^f = .0225$)								
α	T^*	$N(T^*)$	$\mathbf{E}[\mu_{t+1} - \mu_t]$	$\mathcal{T}_{\min}$	$N(\mathcal{T}_{\min})$	R&D	FC	Π
.20	12	19	.211	6	8	2.76	8.55	.32
.35	6	20	.383	3	5	2.92	9.00	.36
.50	4	19	.525	2	4	2.91	8.55	.49

Medium Entry Cost ($x^f = .045$)								
α	T^*	$N(T^*)$	$\mathbf{E}[\mu_{t+1} - \mu_t]$	$\mathcal{T}_{\min}$	$N(\mathcal{T}_{\min})$	R&D	FC	Π
.20	26	9	.111	8	1	2.43	8.10	.06
.35	12	10	.223	6	4	2.77	9.00	.41
.50	9	11	.320	4	4	3.15	9.90	.38

High Entry Cost ($x^f = .0675$)								
α	T^*	$N(T^*)$	$\mathbf{E}[\mu_{t+1} - \mu_t]$	$\mathcal{T}_{\min}$	$N(\mathcal{T}_{\min})$	R&D	FC	Π
.20	40	5	.070	15	1	1.92	6.75	.04
.35	21	7	.167	7	1	2.76	9.45	.22
.50	12	7	.237	5	2	2.84	9.45	.32

Table B.5: Optimal policy of success probability α