Examination of the Relationship Between Competition and Innovation: Toward a Robust Approach

by

Devin D. Garcia

A Thesis Submitted in Partial Fulfillment of the Requirements for the Degree Doctor of Philosophy

Approved, Thesis Committee:

Robin C. Sickles, Chair
Professor of Economics

Richard Boylan
Professor of Economics

James Weston
Professor of Finance

Houston, Texas
April, 2016
ABSTRACT

Examination of the Relationship Between Competition and Innovation:
Toward a Robust Approach

by

Devin D. Garcia

Interest in the relationship between competition and innovation has seen a resurgence in the last decade. Driven by the theoretical possibility of an inverted-U relationship, current research has focused on non-linear models of competition and innovation. The empirical results that proceed from this research are mixed, including predictions of an inverted-U, a monotonically increasing and a monotonically decreasing relationship.

While much attention has been given to the theoretical possibility of a non-linear relationship, relatively little has been given to the subject of measurement. Following Carl Shapiro (2012), I define “more competitive” as the extent to which a firm stands to lose profitable sales to its rivals should it offer inferior value to consumers. My framework implements this definition in a direct way: two firms must simultaneously choose their innovation strategies under the expectation that, should only one successfully innovate, the unsuccessful firm will have a portion of its sales stolen by its rival. The greater the portion, the more contestable, and therefore competitive, the market. This framework predicts a robust, positive relationship.

I apply my model to a sample of U.S. publicly traded manufacturing firms over the period 1962-2009. Innovation is measured via total factor productivity, and compe-
tition is measured as the elasticity of firm market value with respect to sales, where sales proxy consumer value. My measure of innovation is consistent with the fact that innovation drives long-run economic growth, and my empirical measure of competition is consistent with “more competitive” in that it estimates how much market value a firm would have lost if it hypothetically generated less value for consumers. I estimate a dynamic panel model at the firm level with a quadratic specification in competition. My results indicate a positive and monotonically increasing relationship between competition and innovation.
Contents

Abstract i
List of Illustrations vi
List of Tables vii

1 Competition and Innovation:

A Review of the Literature 1

1.1 Preface .................................................. 1
1.2 Introduction ........................................... 4
1.3 Early IO Literature: Market Structure, Incumbency and the Incentive to Innovate .................................................. 15
  1.3.1 Theoretical models .................................. 15
  1.3.2 Empirical models ................................... 19
1.4 New IO Literature: Competition as a Parameter and the Incentive to Innovate .................................................. 20
  1.4.1 Theoretical models .................................. 20
  1.4.2 Empirical models ................................... 31
1.5 Related Empirical Literature Outside IO ......................... 37
  1.5.1 Using natural experiments to identify the effect of competition .................................................. 37
1.6 Problem Statement ...................................... 40

2 Competition and Innovation:

A Theoretical Framework 45

2.1 Introduction ........................................... 45
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.2</td>
<td>Market Power</td>
<td>46</td>
</tr>
<tr>
<td>2.3</td>
<td>Traditional Measures of Competition</td>
<td>52</td>
</tr>
<tr>
<td>2.3.1</td>
<td>Market Concentration</td>
<td>52</td>
</tr>
<tr>
<td>2.3.2</td>
<td>Number of Firms and Barriers to Entry</td>
<td>55</td>
</tr>
<tr>
<td>2.3.3</td>
<td>Product Substitutability</td>
<td>56</td>
</tr>
<tr>
<td>2.4</td>
<td>Measuring Innovation</td>
<td>61</td>
</tr>
<tr>
<td>2.5</td>
<td>Toward a Robust Measure of Competition and its Relationship with Innovation</td>
<td>65</td>
</tr>
<tr>
<td>2.5.1</td>
<td>Model 1: Stochastic innovation with a linear penalty</td>
<td>66</td>
</tr>
<tr>
<td>2.5.2</td>
<td>Model 2: Stochastic innovation with a nonlinear penalty</td>
<td>75</td>
</tr>
<tr>
<td>2.5.3</td>
<td>Model 3: Stochastic process innovation in a two-stage game of strategic R&amp;D and price competition with cost-asymmetric firms</td>
<td>84</td>
</tr>
<tr>
<td>2.5.4</td>
<td>Model 4a: Deterministic process innovation in a two-stage game of strategic R&amp;D and price competition with cost-symmetric firms</td>
<td>90</td>
</tr>
<tr>
<td>2.5.5</td>
<td>Model 4b: Deterministic process innovation in a two-stage game of strategic R&amp;D and price competition with cost-asymmetric firms</td>
<td>93</td>
</tr>
<tr>
<td>2.5.6</td>
<td>Model 5: Deterministic product innovation in a two-stage game of strategic R&amp;D and price competition with cost-asymmetric firms</td>
<td>101</td>
</tr>
<tr>
<td>2.6</td>
<td>Applying the model to an empirical framework</td>
<td>107</td>
</tr>
<tr>
<td>3</td>
<td>Competition and Innovation: An Empirical Test</td>
<td>109</td>
</tr>
<tr>
<td>3.1</td>
<td>Introduction</td>
<td>109</td>
</tr>
<tr>
<td>3.2</td>
<td>The Model</td>
<td>113</td>
</tr>
<tr>
<td>3.2.1</td>
<td>Estimation of Dynamic Panel Models</td>
<td>117</td>
</tr>
</tbody>
</table>
Illustrations

3.1 Marginal impact of competition on innovation: Positive $\phi$ . . . . . . . 139
3.2 Marginal impact of competition on innovation: Statistically significant $\phi$ 142

A.1 Industry-level Plots: Time series . . . . . . . . . . . . . . . . . . . . . . . . . 150
A.2 Summary of Industry Productivity and Competition Calculations . . . 150
A.3 Industry-level Plots: Innovation against competition (standardized) . 151
A.4 Industry-level Plots: Competition against standard measures of
market power and market structure (standardized) . . . . . . . . . . . . . 152
Tables

3.1 Results with positive $\phi$ .................................................. 146
3.2 Results with statistically significant $\phi$ .................................. 147
3.3 Results using different instrument lag lengths .......................... 148
Chapter 1

Competition and Innovation:
A Review of the Literature

1.1 Preface

Innovation, as Joseph Schumpeter put it, is the “fundamental impulse that sets and keeps the capitalist engine in motion.” In other words, innovation is economic progress—and it would be hard to argue otherwise. Look no further than the steam engine, light bulb, automobile, airplane, radio and integrated circuit for clear evidence of this claim. Our understanding of the economics of innovation is, therefore, critical. In my dissertation, I examine how the incentive to innovate is affected by competition.

Economic interest in the relationship between competition and innovation dates back to Schumpeter. He argues that perfect competition is not only inferior to large scale enterprise with respect to innovation, but in fact incompatible with it. Since then, many economists have empirically tested the so-called Schumpeterian hypothesis that market power fosters innovation.

Beginning in the 1960s and ending in the 1980s, focus was mainly on firm size and market concentration. Little was found in the way of a robust or economically significant relationship to innovation. Interest then turned to the relationship between \textit{ex-ante} market power and innovation. Some of these studies find a negative relationship, some a positive and some an inverted-U.
A related but distinct empirical literature presents strong evidence of a positive relationship between competition and innovation. This literature examines changes in productivity at the micro level in response to discrete changes in the competitive environment. That is, this literature uses natural experiments to identify changes in competitive intensity. For example, one study finds that a group of U.S. iron ore manufacturers nearly doubled their labor productivity in the 1980s subsequent to Brazil entering the market with highly competitive prices.

Economic theory also examines the relationship between competition and innovation. Broadly speaking, there are two approaches. One approach compares how much an incumbent firm would be willing to invest in R&D compared to a potential entrant, and the other approach examines the relationship between a parameter that affects market structure and R&D intensity. Like the empirical industrial organization literature, the results are mixed. The incumbent-entrant class of models predict both a positive and negative competition-innovation relationship, while predictions from the parametric class of models range from a positive relationship to an inverted-U.

Overall, both empirical evidence and theory paint an unclear picture of the competition-innovation relationship. One might argue that this is to be expected given the abstract and complex nature of these variables. I make the argument, however, à la Shapiro (2012), that the unclear picture is largely a product of inconsistent and superficial measures of competition and innovation. This is particularly true of competition, where it is standard to define it in terms of market structure and market power. For example, it is standard to associate greater market power, i.e., a greater divergence between price and marginal cost, with “less competition.” But it is not difficult to see that a change in market power, which is essentially a change in profit margin, could identify something other than a change in competitive intensity. Take the life cycle
of a product. For many goods, as a product matures focus shifts from differentiation and quality improvement to cost reduction (Utterback and Abernathy (1975)). In other words, price becomes more important in the later stages of a product’s life cycle. In turn, margins go down. This decrease in margin, however, is not necessarily indicative of a more competitive environment. It merely reflects what often happens to so many goods in the competitive process: commoditization.

Following Shapiro (2012), I define a more competitive industry as a more contestable one, i.e., one where a firm stands to lose a greater amount of profitable sales to its rivals should it offer inferior value to consumers. I apply this definition directly. For example, in a model where the outcome of innovation is stochastic, I examine aggregate R&D intensity when two firms act on the belief that, should one fail to innovate and the other succeed, the unsuccessful firm will have a fraction of its profits captured by its rival. The greater this fraction, the more contestable, and therefore competitive, the market. This is precisely the theme, or question, of my research: how does innovation respond to competition when competition is defined in terms of its power to discipline? My research indicates that the answer to this question is unambiguous: a more competitive environment stimulates innovation.

The rest of my dissertation is organized as follows. In Chapter 1, I survey the theoretical and empirical literature on competition and innovation. In Chapter 2, I develop a theoretical framework where a more competitive market is measured in terms of how contestable it is. In Chapter 3, I develop an empirical measure of competition that identifies how contestable a market is and examine its relationship with productivity, my measure for innovation.
1.2 Introduction

What makes capitalism go, and why is it superior to socialism? This question gave us “Capitalism, Socialism and Democracy,” a book widely viewed as one of the most influential books in economic history. Published in 1942, Joseph Schumpeter wrote “Capitalism, Socialism and Democracy” because he was dissatisfied with the conventional answer to the question posed at the outset; namely, capitalism is superior because the “profit interest of the producer tends to maximize production.” Schumpeter dismissed the idea altogether:

The [conventional argument], as far as it can be proved at all, applies to a state of static equilibrium. Capitalist reality is first and last a process of change. In appraising the performance of competitive enterprise, the question whether it would not tend to maximize production in a perfectly equilibrated stationary condition of the economic process is hence almost, though not quite, irrelevant.

That capitalism is process of constant change is a point made throughout the book:

“... industrial mutation – if I may use that biological term – that incessantly revolutionizes the economic structure from within, incessantly destroying the old one, [and] incessantly creating a new one .. is the essential fact about capitalism ... and [t]he fundamental impulse that sets and keeps the capitalist engine in motion comes from the new consumers’ goods, the new methods of production or transportation, the new markets, [and] the new forms of industrial organization that capitalist enterprise creates.”
These observations led Schumpeter to conclude that innovation, not price competition nor market structure, is what matters for economic growth.

The first thing to go is the traditional conception of the *modus operandi* of competition. Economists are at long last emerging from the stage in which price competition was all they saw. As soon as quality competition and sales effort are admitted into the sacred precincts of theory, the price variable is ousted from its dominant position. However, it is still competition within a rigid pattern of invariant conditions, methods of production and forms of industrial organization in particular, that practically monopolizes attention. But in a capitalist reality as distinguished from its textbook picture, it is not that kind of competition which counts but the competition from the new commodity, the new technology, the new source of supply, the new type of organization (the largest-scale unit of control for instance)—competition which commands a decisive cost or quality advantage and which strikes not at the margins of the profits and the outputs of the existing firms but at their foundations and their very lives.

The prevailing view at the time was that in order to maximize economic welfare, we must strive for markets that approximate the behavior of perfectly competitive ones. The reason: only in perfectly competitive markets, where price meets marginal cost in equilibrium, does the sum of consumer and producer surplus reach its maximum. Of course, every economist knew then as they do now that the assumptions required (*e.g.*, zero barriers to entry and perfect information) to attain marginal cost pricing are very strong. It was the principle that mattered, though, not the end point. Schumpeter
nevertheless rejected it, and not just on the grounds that market structure is irrelevant to economic progress, but that perfect competition itself is incompatible with it.

Perfect competition implies free entry into every industry. It is quite true, within that general theory, that free entry into all industries is a condition for optimal allocation of resources and hence for maximizing output. If our economic world consisted of a number of established industries producing familiar commodities by established and substantially invariant methods and if nothing happened except that additional men and additional savings combine in order to set up new firms of the existing type, then impediments to their entry into any industry they wish to enter would spell loss to the community. But perfectly free entry into a new field may make it impossible to enter it at all. The introduction of new methods of production and new commodities is hardly conceivable with perfect—and perfectly prompt—competition from the start. And this means that the bulk of what we call economic progress is incompatible with it.

Schumpeter was not a proponent of monopoly in the textbook sense. He did advocate for patents, trade secrecy and long-period contracts, all of which facilitate temporary monopoly power, but only because he believed they were necessary to incentivize innovation.

Practically any investment entails, as a necessary complement of entrepreneurial action, certain safeguarding activities such as insuring or hedging. Long-range investing under rapidly changing conditions, especially under conditions that change or may change at any moment under the impact of new commodities and technologies, is like shooting at a tar-
get that is not only indistinct but moving—and moving jerkily at that. Hence it becomes necessary to resort to such protecting devices as patents or temporary secrecy of process or, in some cases, long-period contracts secured in advance.

He also credits “big business” for its innovation and defends practices like excess capacity, but only if such practices are used as a temporary stop-gap to secure a position for the future.

As soon as we go into details and inquire into the individual items in which progress was most conspicuous, the trail leads not to the doors of those firms that operate under conditions of comparatively free competition but precisely to the doors of the large concerns ... and a shocking suspicion dawns upon us that big business may have had more to do with creating [a greater] standard of life than with keeping it down...

The main value to a concern of a single seller position that is secured by a patent or monopolistic strategy does not consist so much in the opportunity to behave temporarily according to the monopolist schema, as in the protection it affords against temporary disorganization of the market and the space it secures for long-range planning.

Overall, Schumpeter’s thesis is a nuanced one. He is clear about one thing, however: in order to make any assessment of economic performance within a capitalist reality, we must recognize that capitalism is first and foremost a system of perennial change. And in recognizing the latter, Schumpeter makes the controversial claim that perfect competition is not only incapable of driving economic growth, but that “big business” is superior to it:
Thus it is not sufficient to argue that because perfect competition is impossible under modern industrial conditions—or because it always has been impossible—the large-scale establishment or unit of control must be accepted as a necessary evil inseparable from the economic progress which it is prevented from sabotaging by the forces inherent in its productive apparatus. What we have got to accept is that it has come to be the most powerful engine of that progress and in particular of the long-run expansion of total output not in spite of, but to a considerable extent through, this strategy which looks so restrictive when viewed in the individual case and from the individual point in time. In this respect, perfect competition is not only impossible but inferior, and has no title to being set up as a model of ideal efficiency. It is hence a mistake to base the theory of government regulation of industry on the principle that big business should be made to work as the respective industry would work in perfect competition.

Schumpeter’s position on big business, perfect competition and innovation is not a statement about the relationship between market power and innovation, however. While he did argue that innovation is incompatible with perfect competition and perhaps buoyed by scale and the prospect of market power, this is not equivalent to the statement that innovation suffers under more intense competition or that it thrives with greater *ex-ante* market power. Nevertheless, Schumpeter is commonly associated with the hypothesis that market power spurs innovation—a hypothesis which ranks as one of the most widely tested in the industrial organization (IO) literature.
A substantial IO literature on competition and innovation has developed for the better part of the last century. Until relatively recently, tests of the Schumpeterian hypothesis were direct in that they examined the relationship between market structure/firm size and innovation. This characterizes what I call the “early” literature. A “new” literature has taken hold, however, where express interest is in the relationship between competition and innovation.

The early literature is defined mainly by incumbent-entrant models and cross-sectional studies on the relationship between R&D and firm size/concentration. Incumbent-entrant models test the Schumpeterian hypothesis by seeing which of the two types of firms has a greater incentive to innovate. If the incumbent is found to have a greater incentive to innovate—defined as the difference between post- and pre-innovation rents—then market power spurs innovation. That is, incumbent-entrant models test the Schumpeterian hypothesis by examining whether or not the possession of market power is a driver or deterrent of innovation. Arrow (1962), Gilbert and Newbery (1982) and Reinganum (1983)—two of which find that the possession of market power deters innovation—define this literature. The empirical literature is much larger. But across these studies, the core relationship tested is that between scale and R&D activity. Scale is measured in basically two ways: firm size and market concentration. Innovation, or more precisely innovative effort, is measured using R&D expenditures or R&D intensity, the ratio of R&D expenditures to sales. Scherer (1967) represents one such paper in this literature. He finds in the cross-section that the fraction of R&D personnel to total personnel initially increases with market concentration, but then declines. In general, however, this literature finds little to no evidence of a statically significant relationship between scale and R&D activity.
The “new” literature is expressly concerned with the relationship between competition—not scale or market structure, per se—and innovation. Moreover, unlike the early literature, the level of competition is treated as a continuum. Take Aghion et al. (2005) (ABBGH henceforth), for example, which has made a substantial impact on the literature. As opposed to early theoretical work that focuses on “discrete” differences in market power and its effect on innovation, ABBGH measure market power as the extent to which two firms, namely a “neck-and-neck” firm and “leader” firm, can collude. That is, they identify a more competitive market as one where a neck-and-neck industry faces lower profits. Coupled with a dynamic model of step-by-step innovation—where successful innovation can narrow or widen the technology gap between rivals by one step—they find that competition and innovation share an inverted-U relationship. They confirm their theoretical result using data on a panel of publicly traded manufacturing firms in the UK, where innovation is measured using citation-weighted patent counts and competition is measured using a proxy for the Lerner index.

Since ABBGH, emphasis has been on empirical work. Three derivative papers, in fact, test the ABBGH model using similar data and methods: Correa (2012), Correa and Ornaghi (2014) and Hashmi (2013). Despite the similarities, the results vary qualitatively. Correa (2012), who uses the same data and model as ABBGH, finds that the inverted-U breaks down if a structural break in the data is accounted for. In particular, he finds a positive relationship before the structural break, but then a non-statistically significant one thereafter. Correa and Ornaghi (2014) and Hashmi (2013) also arrive at different results. Correa and Ornaghi (2014) finds evidence of a monotonically positive relationship, whereas Hashmi (2013) finds a negative one.
The difference in results is somewhat surprising given the similar methods and data employed; this alone warrants further investigation of the relationship between competition and innovation. More importantly, however, is the lack of attention that has been given to measurement. Consider first the measurement of competition. A standard measure of competition in the literature is the Lerner index, or the difference between a firm’s price and its marginal cost, normalized by price. The popularity of this measure stems from its relationship with the price elasticity of demand, a classical source of market power. Indeed, when a single, profit maximizing firm takes its demand as given, its equilibrium Lerner index will be inversely proportional to its price elasticity of demand. Thus, a larger Lerner index is taken to indicate greater market power or, as it is often interpreted, less competition. There are several reasons why this measure could fail to indicate changes in market power, however. One possibility, for example, is if a firm engages in cost-minimizing behavior. In this case, a temporal increase in the Lerner index would reflect nothing more than the competitive process at work.

Other common measures of competition include the Herfindahl-Hirschman index, the number of firms in an industry, market share and, at a deeper theoretical level, the degree of product substitutability. Each of these have some relationship with market structure. For example, the Herfindahl-Hirschman index corresponds to monopoly when the market is captured entirely by one firm and, on the other end, perfect competition when there are infinitely many firms with equal shares. Every measure of competition used in the literature is thus based on the presumption that changes in market structure or market power will identify changes in competitive intensity. However, this presumption is questionable if we identify a more competitive market as one with greater rivalry to provide a superior product or service. It does not follow
that a more profitable firm or a more concentrated industry will be less rivalrous, especially when changes in market structure or market power are small. A more competitive market will arise when, on average, a firm faces a greater risk of falling behind should it offer inferior value to consumers. That is, a more contestable market will result in greater rivalry. Shapiro (2012) calls this the “contestability” principle, which I adopt to examine the relationship between competition and innovation.

Application of the “contestability” principle is consistent with antitrust policy. The 2010 Horizontal Merger Guidelines of United States Department of Justice and the Federal Trade Commission state the following:

The unifying theme of these Guidelines is that mergers should not be permitted to create, enhance, or entrench market power or to facilitate its exercise... [where by definition] a merger enhances market power if it is likely to encourage one or more firms to raise price, reduce output, diminish innovation, or otherwise harm customers as a result of diminished competitive constraints or incentives.

Notice that the Guidelines identify the root of enhanced market power as any market characteristics that would lessen rivalry, or the incentive to compete. While it is possible that enhanced market power could manifest as greater profitability or market concentration, it need not be the case that the reverse hold, i.e., that the latter result

---

1 Anne Bingaman, a former Assistant Attorney General for the Anititrust Division in the U.S. Department of Justice during 1993-1996, had this to say about rivalry and innovation: “The fundamental thesis of strong antitrust enforcement is that rivalry, not market power, fosters innovation and efficiency over the long run... Antitrust has an important role in preserving the rivalry that spurs innovation.” See Bingaman (1994).

2 See HMG (2010) for the full set of guidelines.
in or correspond to enhanced market power. On the other hand, any policy that would serve to make a market more contestable will, all else equal, intensify competition.

The measurement of innovation is equally important. By definition, an innovation is a new and/or significantly improved product, process, service or business method. Thus, at least in a stylized theoretical model, innovation must satisfy three criteria: it generates value, requires investment and, optionally, is the result of a stochastic process. For example, Aghion et al. (2005) measure innovation in terms of R&D intensity, where greater R&D intensity increases the chance of successful innovation.

I do not deviate from this stylized operationalization of innovation in a theoretical setting. However, regarding empirical measures, I eschew the recent trend toward citation-weighted patent counts as a measure of innovation in favor of productivity.

A citation-weighted patent count is as it sounds: the number of citations a patent has. There are several problems with patents as a measure of innovation, however. One, patents represent only a fraction of innovative output. Two, the incentive to patent goes beyond the intent to protect a novel and valuable idea; companies also seek out patents to defend against litigation or to launch a lawsuit themselves. Three, evidence suggests that companies view patents as relatively weak mechanisms for protecting intellectual assets; instead, trade secrecy and leadtime to market are viewed as more effective. And four, the presumption that a greater number of citations reflects a more valuable idea has been challenged. Empirical evidence indicates that the relationship between citations and economic value is an inverted-U. Intuitively, the most valuable patents are the ones firms actively try to protect the most, resulting in fewer citations.

For the reasons above, I use productivity as my measure of innovation. Productivity growth, as Jorgenson (2011) articulates well, is the key economic indicator of
innovation:

Productivity growth is the key economic indicator of innovation. Economic growth can take place without innovation through replication of established technologies. Investment increases the availability of these technologies, while the labor force expands as population grows. With only replication and without innovation, output will increase in proportion to capital and labor inputs, as suggested by Schultz (1956, 1962). By contrast the successful introduction of new products and new or altered processes, organization structures, systems, and business models generates growth of output that exceeds the growth of capital and labor inputs. This results in growth in multifactor productivity or output per unit of input.

Thus, not only does productivity as a measure avoid the economic complications of patents, it is consistent with long-run economic growth—precisely why Schumpeter was such an advocate for anything that would serve to promote innovation.

The disconnect between theory and practice is not uncommon. In the case of IO models of competition and innovation, the disconnect primarily emanates from the literature’s equivalence of “more competitive” to superficial notions of greater market power, which may or may not reflect how intense competition is in any particular market. Ultimately, what makes competition a powerful force for innovation is its ability to discipline. Companies that face little threat of falling behind even if they offer an inferior product or service to consumers will inevitably be less inclined to innovate. This point is the essence of my research, which is supported by a rigorous theoretical and empirical investigation of the relationship between contestability and innovation. The literature review now follows.
1.3 Early IO Literature: Market Structure, Incumbency and the Incentive to Innovate

1.3.1 Theoretical models

The early theoretical literature is characterized by a focus on market structure and incumbency and their effect on the incentive to innovate—defined as the difference be post- and pre-innovation rents. Arrow (1962) was the first to rigorously examine the relationship between market structure and innovation. He demonstrates that a perfectly competitive firm has more to gain from innovation than a monopolist when the monopolist faces no threat of entry and when innovation is cost-reducing. He reasons that this is the case because, unlike a perfectly competitive firm, the monopolist must “replace” itself; i.e., a perfectly competitive firm has no profit stream to replace, whereas the monopolist does.

Two other papers build off of the Arrow model. Gilbert and Newbery (1982) relax the assumption of no entry into the monopolist’s market, and Reinganum (1983), in addition to allowing for entry, relaxes the assumption of deterministic innovation. The relaxation of no entry gives the model a strategic element. Now two firms, an incumbent and challenger, compete via innovation to secure a patent which grants an exclusive and perpetual right to the innovator’s technology. In the case of deterministic innovation, Gilbert and Newbery find that the incumbent has a greater incentive to innovate because entry generates a larger loss for the incumbent than a gain for the challenger. However, when innovation is stochastic, Reinganum finds that the challenger has a greater incentive to innovate because the marginal benefit to the incumbent from investing marginally less in R&D is greater than that of the challenger. A review of the above models now follow.
Review of the literature

The relationship between market structure and innovation starts with Kenneth Arrow’s seminal paper, “Economic Welfare and the Allocation of Resources for Invention.” Arrow (1962) shows that a monopolist with no threat of entry has a lower incentive to innovate than a perfectly competitive firm.

Arrow models innovation as an investment that serves to lower a firm’s marginal production cost. Marginal cost is assumed to be constant and equal to $c_0$ for an old technology and $c_1 < c_0$ for a new technology. In principle, the incentive to innovate is the same for both a monopolist and a competitive firm: profit with the new technology less the profit with the old technology. The main difference is that a competitive firm collects zero *ex-ante* profits.

There are two types of innovations: drastic and non-drastic. A drastic innovation is one where the profit maximizing price corresponding to the new technology is less than the marginal cost of the old technology, and a non-drastic innovation is otherwise. Thus, when innovation is drastic, the post-innovation rent will be the same for a competitive firm and a monopolist; i.e., both firms will earn a monopoly profit. As a result, the incentive to innovate will be higher for a competitive firm because its pre-innovation rent of zero is lower than a monopolist’s.

In the event of non-drastic innovation, the post-innovation rent for a competitive firm will be less than a monopolist’s. This is because a competitive firm cannot charge a price that exceeds the prevailing price, $c_0$, in a competitive market where the old technology remains.\(^3\) Even so, Arrow shows that a monopolist’s net payoff will be less

---

\(^3\)A competitive firm’s post-innovation rent is therefore the unit royalty it earns ($c_0 - c_1$) times the number of units sold in the market at price $c_0$. This assumes an innovator can secure a patent for its innovation.
than a competitive firm’s precisely because the monopolist’s marginal revenue curve is decreasing. Thus, regardless if innovation is drastic or non-drastic, a competitive firm has a higher incentive to innovate than a monopolist.

Gilbert and Newbery (1982) relax Arrow’s model by allowing for the possibility of entry by a challenger into the monopolist’s market. This gives the model a strategic component. Specifically, an incumbent and a potential entrant now bid for a patent on a new, substitute technology that would serve to preempt entry from the perspective of the incumbent or enable it from the perspective of the potential entrant. If the incumbent wins, it maintains its monopoly. If the challenger wins, it enters the market and competes with the incumbent.

What they find is that preemption is a rational strategy—in fact, a Nash equilibrium—for the incumbent if monopoly profits with the new technology exceed the costs of preemption. This will attain if post-entry industry profit is less than pre-emptive monopoly profit. That is, letting \( m \) denote monopolist and \( e \) entrant, if

\[
\pi_m(p^1_m, p^2_m) > \pi_m(p^1_m, p^2_e) + \pi_e(p^1_m, p^2_e),
\]

(1.1)

attains, where \( \pi_i(\cdot), i \in \{m, e\} \), represents profit and \( p^j_i, j \in \{1, 2\} \), represents the price of product \( j \). This will in fact attain under very weak assumptions.

Note now that in their model \( \pi_e(p^1_m, p^2_m) = 0 \). Thus, the incumbent will have an incentive to preemptively innovate if and only if its incentive to innovate is greater than that of the entrant’s. This is easily seen by rearranging (1.1):

\[
\pi_m(p^1_m, p^2_m) > \pi_m(p^1_m, p^2_e) + \pi_e(p^1_m, p^2_e) > \pi_e(p^1_m, p^2_e).
\]

\(^4\)Here, product 2 is the new, substitute technology being auctioned.

\(^5\)This assumption does not change their result.
In other words, their model shows that despite a monopolist having to replace itself, a monopolist may still be more inclined than an entrant to innovate because entry would substantially reduce its profits.

Reinganum (1983) extends the model of Gilbert and Newbery (1982) by making the date of successful innovation stochastic. Specifically, she assumes that $\Pr(T(x) \leq t) = 1 - e^{-h(x)t}$, where $T(\cdot)$, $h(\cdot)$ and $x$ are the random date of successful innovation, hazard rate of successful innovation and R&D expenditure, respectively. Her model now follows.

An incumbent and challenger invest $x_I$ and $x_C$, respectively, in R&D to reduce their marginal cost of production. If the incumbent succeeds in reducing its current marginal cost from $\bar{c}$ to $c < \bar{c}$ and secures a patent before the challenger, it will earn flow profit $\Pi(c)$. If the challenger succeeds before the incumbent, the incumbent will earn $\pi_I(c) < \Pi(c)$ and the challenger $\pi_C(c)$. If neither succeed, the incumbent and challenger maintain their pre-innovation profit flows $R$ and zero, respectively.

It is assumed that $\Pi(c)$ and $\pi_C(c)$ are non-increasing and $\pi_I(c)$ non-decreasing in $c$, the intuition being that the successful (unsuccessful) innovator’s rents are higher (lower) the greater is its reduction in cost. Moreover, the industry is characterized by constant returns to scale, which implies full capture of the market should the challenger generate a drastic innovation.

A drastic innovation is formally defined as an innovation which lowers marginal cost to $c \leq c_0$, where $c_0$ is assumed to exist and is the maximum value of $c$ such that $\pi_I(c) = 0$. The significance of the constant returns to scale assumption is that zero profits implies zero output, so in the event of a drastic innovation, $\pi_C(c) = \Pi(c)$ (i.e., the challenger becomes a monopolist).

The expected future discounted profit for the incumbent and challenger are given
by, respectively,

\[ V^I(x_I, x_C) = \int_0^\infty e^{-rt}e^{-(h(x_C)+h(x_I))t}[h(x_I)\Pi(c) + h(x_C)\pi_I(c) + R - x_I]dt \]

\[ V^C(x_I, x_C) = \int_0^\infty e^{-rt}e^{-(h(x_C)+h(x_I))t}[h(x_C)\pi_C(c) - x_C]dt, \]

where \( r \) is the interest rate.

Reinganum implicitly characterizes the optimal Nash R&D levels of this game. She proves that for drastic innovations, a challenger will spend more on R&D than the incumbent. She also shows this to be the case for non-drastic innovations near \( c_0 \), invoking continuity of the Nash equilibrium. Thus, in total, she finds a non-trivial set of innovations where the challenger invests a strictly greater amount in R&D than the incumbent.

1.3.2 Empirical models

The early empirical literature is substantial in size. I focus on one paper: Scherer (1967).

Review of the literature

Scherer (1967) extends the analysis of Scherer (1965) by using a new and “more comprehensive” dataset, which consists of 56 industries. Unlike his previous analysis where firm size was of primary interest, he focuses on market concentration. Moreover, instead of using patents as his proxy for innovation, his primary measure of innovation

\[ \text{Technically, these expressions are approximations because of the way the events are defined. For example, the incumbent will earn } \Pi(c) \text{ at time } t \text{ if the challenger does not succeed by time } t \text{ and the incumbent succeeds precisely at } t. \text{ The latter event has no mass.} \]

\[ \text{See Gilbert (2006) for a full survey.} \]
is the total number of engineers and scientists employed as a proportion of total employees. Thus, he moves from an output-based to an input-based measure.

There are two findings that stand out in Scherer’s cross-sectional study. One is that the explanatory power of market concentration drops substantially once industry-specific technological opportunity is controlled for. The other is that innovation effort appears to exhibit an inverted-U relationship with concentration. The former result is a common feature of the early empirical literature. And the latter result predates what is now standard today: non-linear competition-innovation specifications.

1.4 New IO Literature: Competition as a Parameter and the Incentive to Innovate

1.4.1 Theoretical models

Early theoretical models may be viewed as “discrete” in the sense that they compare the incentive to innovate across two types of firms: one with market power and one without. “New” theory, however, models the degree of competition as a continuum. Five such models are presented.

Kamien and Schwartz (1976) were the first to find a theoretically non-linear relationship between competition and innovation. In their framework, a firm chooses its introduction date of innovation based on, among other parameters, the degree of rivalry it faces. The extent of rivalry is measured in terms of the expected date of innovation by a rival. They find an inverted-U relationship between speed of innovation and rivalry.

The second model, Loury (1976), builds off Kamien and Schwartz. Unlike Kamien and Schwartz, however, Loury allows for interdependency between the actions of
rivals. That is, he examines how R&D incentives are affected by rivalry when rivals choose their strategies simultaneously. For tractability, Loury imposes the assumption of symmetry across firms and consequently measures rivalry in terms of the number of firms in the industry. These modifications generate the prediction that greater rivalry will lead to less R&D expenditure in aggregate but an earlier expected date of innovation.

Breaking away from the static models of Kamien and Schwartz and Loury, Aghion et al. (2001), or AHHV for short, develop a dynamic model of “step-by-step” innovation. By “step-by-step,” AHHV mean the technological gap between rivals can be narrowed or widened step-by-step through innovation. For example, in their model which consists of two firms, a “leader” firm will increase its lead by one step if it successfully reduces its cost of production and the “laggard” fails. How competition affects this process is conceptually straightforward. Two firms, asymmetric in terms of their marginal cost of production, price compete with a differentiated good. The equilibrium profit function from this stage game—which implicitly depends on relative efficiency and the degree of product substitutability—is the underlying payoff function in their step-by-step model. The solution to their dynamic model is therefore an implicit function of product substitutability, or their measure of competition. They find the following. When innovation is very large, i.e., a one step lead amounts to the leader capturing the entire market, competition initially stimulates innovation but then thwarts it. For all other size-variations of innovation, they find a monotonically positive relationship.

Aghion et al. (2005), or ABBGH, modify the model of AHHV by no longer assuming a laggard can immediately catch to the leader if it innovates. They restrict attention to the case where the maximum sustainable technological gap between leader
and laggard is one, which implies that the leader will never innovate. Like AHHV, they find an inverted-U relationship between competition and innovation when the size of innovation is large.

Finally, Hashmi (2013) revisits the model of ABBGH and finds that the inverted-U relationship is a special case. More specifically, he shows via simulation that the competition-innovation relationship depends on the technological gap in an industry. When the gap is relatively small, innovation increases with competition. When the gap is intermediate, the relationship is an inverted-U. And when the gap is large, the relationship is negative. His results are based on a partial equilibrium analysis of ABBGH.

**Review of the literature**

Kamien and Schwartz (1976) develop two frameworks to analyze the effect of rivalry on the speed of innovation. In both frameworks, they find that the introduction date of innovation is minimized at an intermediate level of rivalry. That is, the relationship between the speed of innovation and rivalry is U-shaped. I give an overview of one of these frameworks.

Kamien and Schwartz attack the problem from the perspective of a firm seeking to maximize its present expected value of rewards from innovation given the possibility of a rival innovating first. That is, the firm chooses its development date, \( T \), to solve

\[
\max_T \int_T^\infty e^{-(r-g)t} [P_0(1 - F(t)) + P_1(F(t) - F(T)) + P_2F(T)]dt - C(T),
\]

where \( r > g \) is the discount rate, \( g \) is market growth (which may be positive or negative), \( F(\tau) \) is the probability of rival introduction by time \( \tau \), \( P_0(1 - F(t)) \) is the expected payoff to the firm conditional on no rival entry by time \( t \), \( P_1(F(t) - F(T)) \) is
the expected payoff to the firm conditional on it innovating before a rival and the rival appearing between time $T$ and $t$, $P_2$ is the expected payoff to the firm conditional on a rival innovating before time $T$ and $C(T)$ is the minimum present value of the cost of completing development by time $T$. It is assumed that $C'(T) < 0$ and $C''(T) > 0$, and $P_0 \geq P_1$ and $P_0 \geq P_2$.

Once the model is solved, Kamien and Schwartz perform a comparative static analysis on $T^*$, the optimal development date. This optimal date is a function of $h \equiv F'(t)/(1 - F(t))$, or the hazard rate of successful innovation by a rival. Kamien and Schwartz assume the probability of innovation is memoryless, meaning the hazard rate is constant and $F(\tau) = 1 - e^{h\tau}$ if $\tau \in [0, T]$.

The significance of $h$ is its interpretation as a measure of rivalry. Because the inverse of $h$ is the expected date of innovation by a rival, $h$ may be viewed as a proxy for the intensity of rivalry. Kamien and Schwartz adopt this interpretation. Taken as such, they find that $T^*$ and $h$ share a U-shape relationship. Thus, optimal development time initially declines as rivalry intensifies, reaches a minimum, then increases.

The intuition is straightforward. The firm is always weighing the marginal cost of postponement against the marginal loss of delay; if the marginal cost of postponement is greater than the marginal loss of delay, then the firm will hasten its development date. Consider the case of $h = 0$. When $h = 0$, there is no threat of rivalry, meaning the marginal cost of postponement is very small. This will result in a very late development date. As soon as $h$ turn positive, however, the threat of rivalry adds to the marginal cost of postponement. In particular, earlier innovation by a rival results in foregone quasi rents to the firm. This incentivizes the firm to choose an earlier innovation date. Finally, while increases in $h$ add to the marginal cost of postpone-
ment, the firm must weigh this against the increasing cost of an earlier innovation date. Indeed, the latter cost eventually dominates the former, so \( T^* \) increases past some threshold value of \( h \).

Loury (1976) adds to the model of Kamien and Schwartz by allowing for strategic interaction among a finite number of symmetric firms. He finds that the U-shaped result in Kamien and Schwartz is not robust. Moreover, he finds that greater rivalry, measured as a higher number of firms, lowers individual R&D expenditures, but simultaneously accelerates the introduction of innovations.

Like in Kamien and Schwartz, the date of successful innovation is assumed to be memoryless. Thus, the probability of successful innovation by time \( t \) is given by:

\[
P(\tau(x) \leq t) = 1 - e^{-h(x)t},
\]

where \( \tau(\cdot) \) is the random date of successful innovation, \( x \) is R&D expenditure and \( h(\cdot) \) is the hazard rate of innovation. It is assumed that \( h'(x) > 0, h''(x) \geq 0 \) for \( x \in [0, \bar{x}] \) and \( h''(x) < 0 \) for \( x \in (\bar{x}, \infty) \).

For any firm \( i \), the probability that it successfully innovates before any of its rivals is

\[
P(\tau(x_i) \leq t) = 1 - e^{-a_i t},
\]

where \( a_i \equiv \sum_{j \neq i} h(x_j) \). Thus, letting \( V \) denote the revenue flow to firm \( i \) should it innovate before its rivals and \( r \) the interest rate, firm \( i \) solves

\[
\max_x \frac{Vh(x)}{r(a_i + r + h(x))} - x.
\]

The symmetry of the problem implicitly defines the optimal R&D expenditure function

\[
x^* = \hat{x}((n - 1)h(x^*), r, V).
\]
From this, Loury shows that \( \partial \hat{x}/\partial a \)—the analogue to the comparative static analysis of Kamien and Schwartz—can, depending on the magnitude of \( h(\hat{x}) \), be negative everywhere or initially positive then negative. That is, the result of Kamien and Schwartz is not robust. More importantly, Loury shows that \( \partial x^*/\partial n < 0 \) and \( \partial \mathbb{E}[\tau(n)]/\partial n < 0 \) for all parameter values. In words, while greater rivalry lowers individual R&D expenditures, it also accelerates the introduction of innovation. Intuitively, an increase in the number of firms in a symmetric industry will lower expected profits, giving firms less of an incentive to invest in R&D. At the same time, however, a increase in the number of firms means more firms are working the same problem, so to speak, and this increases the chance of earlier innovation.

AHHV investigate the effect of competition on innovation in a dynamic model of step-by-step innovation. Deviating from Schumpeterian models of endogenous growth—where the incentive to innovate is defined only in terms of post-innovation rents—they show that competition has an overall positive effect on innovation when the incentive to innovate is instead defined as the innovator’s increment in rents. Specifically, when the incentive to innovate is defined as the difference between post- and pre-innovation rents, they show in a model of economy-wide duopolistic competition that an increase in competition will lead to greater innovation if innovation is not too large. If, however, the size of innovation is large, they show it is possible for innovation and competition to share an inverted-U relationship.

Their model begins with a stage game of differentiated Bertrand competition. There are a continuum of industries, each composed of two cost-asymmetric firms that price compete with differentiated goods. Under the assumption that consumers have log preferences and industry output is a symmetric, homogeneous of degree one function, they derive equilibrium profits as an implicit function of relative cost and
the degree of product substitutability, the latter being their proxy for competition.

Given the stage game, the basic problem a firm faces is to choose a level of R&D intensity that maximizes its expected future discounted profits conditional on the state of the gap in its industry. AHHV analyze this problem under the following assumptions: (1) a steady state composition of \( n \)-gap (unleveled) and 0-gap (leveled) industries; (2) immediate catch-up by the laggard if only the laggard innovates; (3) an at-most one step increase in the gap if a leader or neck-and-neck firm innovates and; (4) changes in competition affect the entire economy. They present their results under three scenarios. They first investigate the equilibrium effect of competition on innovation when innovation size is large. Then they repeat this investigation for small innovations. Finally, they numerically analyze the problem in its general form.

Under the first scenario of large innovation, they find that an increase in competition can either increase or reduce innovation depending on the current level of competition. To see this, first note that large innovation implies that a one-step lead will raise the would-be leader’s profit to the maximal level. This means we can restrict attention to the R&D intensities of neck-and-neck and laggard firms as the leader will not innovate. Second, note that it effectively fixes the post-innovation rent for a neck-and-neck firm and the pre-innovation rent for a laggard. The effect of competition on innovation thus operates only through the pre- and post-innovation rents of neck-and-neck and laggard firms, respectively, when innovation is large. In fact, these rents are the same: the pre-innovation rent for a neck-and-neck firm is the post-innovation rent for a laggard. And since competition lowers neck-and-neck rents, competition will have a non-linear effect on economy-wide innovation.

The inverted-U arises because the steady-state allocation of leveled and unleveled industries depends on the state of competition in the economy. More specifically,
when competition is already relatively high, neck-and-neck firms will have a greater incentive to innovate than laggards as they seek to “escape competition.” This will push the economy into a state where there are more unleveled industries than leveled ones. And because this transition does not affect the state of competition (i.e., the level of neck-and-neck profits), the incentive to innovate across the economy will diminish—laggards have little incentive to innovate when competition is already high, and leaders do not innovate at all. AHHV call this the “Schumpeterian” effect of competition. Contrast this to the case of relatively low competition. When competition is relatively low, laggard firms will be more inclined than neck-and-neck firms to innovate. This will push the economy into a state with more leveled than unleveled industries where the “escape competition” effect dominates. Thus, innovation initially increases with competition but then declines.

In the case of small innovation, innovation is found to monotonically increase with competition. Intuitively, when the size of innovation is small, the increment in profit from innovating is approximately the same for leaders, laggards and neck-and-neck firms; and since the increment in profit for a neck-and-neck firm increases with competition (due to the “escape competition” effect), it follows that economy-wide innovation will simultaneously increase.

Finally, AHHV examine the general case numerically and confirm their analytical results. That is, they find innovation to numerically increase with competition for intermediate values of innovation size, but eventually an inverted-U arises when innovation is large and the probability of imitation is low.

ABBGH is a derivative of AHHV, the difference being that the laggard is no longer capable of immediately catching up with the leader should the laggard innovate. Even though the ABBGH and AHHV models are similar, it is worth going over some of
the details in ABBGH as this paper has had a substantial impact on the literature.

ABBGH reexamine the “step-by-step” innovation model developed in AHHV where, in an economy comprised of a continuum of intermediate two-firm sectors, a firm can advance its technological position by exactly one step through successful innovation. If, for example, the current state of a sector is $m$ — a non-negative integer that indexes the efficiency gap between two firms — and the leader (laggard) successfully innovates while the laggard (leader) does not, then the state of the sector will change from $m$ to $m + 1$ ($m - 1$). Thus, unlike AHHV, this model assumes that the laggard cannot immediately catch up to the leader.

At any point in time, there exists two kinds of intermediate sectors: (i) leveled (i.e., $m = 0$), or “neck-and-neck,” and (ii) unleveled (i.e., $m > 0$). For tractability, ABBGH restrict $m \in \{0, 1\}$, meaning that a leader (laggard) can be at most one step ahead (behind). The restriction on the gap follows, as ABBGH assume, if spillovers prevent the leader from advancing further. This implies that the aggregate innovation rate will depend only on the research intensities of a laggard and neck-and-neck firm.

On innovation, ABBGH assume a Poisson process. Research intensity is accordingly measured as a Poisson hazard rate in their model — the more effort a firm exerts, the higher is its chance of successful innovation. The cost of effort, $n$, is $c(n) = n^2/2$.

Competition is measured as the degree to which neck-and-neck firms cannot collude. Formally, let $\epsilon \in [0, 1/2]$ denote a neck-and-neck firm’s profit as a fraction of a would-be leader’s profit; then smaller $\epsilon$ corresponds to greater competition. Competition is then parameterized as $\Delta \equiv 1 - \epsilon = (\pi_1 - \pi_0)/\pi_1$, where $\pi_0$ and $\pi_1$ are the profit levels of a neck-and-neck and leader firm, respectively.

Two things are of note. First, given the assumption of a maximum technological gap of one, changes in competition only affect neck-and-neck profits. The incentive to
innovate for a neck-and-neck firm is therefore only affected by competition through a reduction in its pre-innovation rent, and for a laggard, only through a reduction in its post-innovation rent. Second, because changes in competition happen at the economy level, the incentive to innovate simultaneously changes for leveled and unleveled industries.

To determine the effect of competition on innovation, ABBGH examine the steady-state of this economy and find an inverted-U relationship. Their intuition is the following. When competition is initially low (i.e., \( \pi_0 \) is high), neck-and-neck industries will be slow to transition to an unleveled state, and unleveled sectors will be quick to transition to a neck-and-neck one. This means that the economy will spend most of its time in a neck-and-neck state where the “escape competition” effect dominates (neck-and-neck research intensity increases with an increase in competition). Consequently, an increase in competition will lead to faster-than-average innovation. On the other hand, when innovation is initially high (i.e., \( \pi_0 \) is low), neck-and-neck sectors will be quick to transition to an unleveled state, and unleveled sectors will be slow to transition to a neck-and-neck one. This means that the economy will spend most of its time in an unleveled state where the “Schumpeterian” effect dominates ( laggard research intensity decreases with an increase in competition). Consequently, an increase in competition will lead to slower-than-average innovation. The end-result is an inverted-U relationship between competition and innovation.

Hashmi (2013) makes the argument that the theoretical model of ABBGH is not well-suited for industry level analysis. In particular, because ABBGH model competition and innovation at the economy level, their results are not testable at the industry level. However, as Hashmi points out, the theoretical foundation of ABBGH is a partial equilibrium model, which may be used to develop an industry-
level framework. Doing so, he finds that the relationship between competition and innovation depends on the technological gap in an industry.

The model starts with a stage game (in fact, the same stage game as in ABBGH). Two cost-asymmetric firms price compete with a differentiated good. The demand for firm $i$, $i = 1, 2$, is given by

$$q_i = \frac{\frac{1}{p_i^{\alpha_i-1}}}{p_i^{\alpha_i-1} + p_{-i}^{\alpha_{-i}-1}},$$

where $p$ is price and $\alpha \in [0, 1]$ is the degree of product substitutability. The latter captures the degree of competition. Given their demand schedules, both firm’s maximize their profit given the price of their rival. The first-order condition for firm $i$ is

$$(p_i - c_i) \frac{\partial q_i}{\partial p_i} + q_i = 0,$$

where $c_i = w^{\gamma^{-k_i}}$ is firm $i$’s constant marginal cost of production; $w$ is the wage rate, $\gamma$ is the size of innovation and $k_i$ is the technology level of firm $i$. Solving the first-order conditions, equilibrium profit for firm $i$ is

$$\pi_i(n) = \frac{(1 - \alpha) R_i(n)}{1 - \alpha R_i(n)},$$

where $n \equiv k_i - k_{-i}$ is the technology gap between firm $i$ and $-i$, and $R_i(\cdot)$ is firm $i$’s market share.

The Bellman equation of this model is given by:

$$V(n) = \max_{x(n) \in (-n)} \{\pi(n) - wx(n)$$

$$+ \beta[P(x(n), n)[1 - P(x(-n), -n)]V(n + 1)$$

$$+ [1 - P(x(n), n)]P(x(-n), -n)V(n - 1)$$

$$+ P(x(n), n)P(x(-n), -n)V(n)$$

$$+ [1 - P(x(n), n)][1 - P(x(-n), -n)]V(n)]\},$$
where $\beta$ is the discount rate, $x(\cdot)$ is R&D intensity, $\pi(\cdot)$ is current profit (derived above), $P(\cdot)$ is the probability of successful innovation and $V(\cdot)$ is the value function.

The second line corresponds to the event of successful innovation by the leader and unsuccessful innovation by the laggard, which increases the technological gap to $n + 1$. The third corresponds to the event of successful innovation by the laggard and unsuccessful innovation by the leader, leading to a technological gap of $n - 1$. And the fourth and fifth lines correspond to, respectively, both firms successfully and unsuccessfully innovating, keeping the technological gap at $n$. The Bellman equation in ABBGH takes a similar form.

The probability of successful innovation is given the following form:

$$P(x(n), n) = [1 - e^{-ax}] + \max\{0, 1 - e^{\eta(n - \hat{n})}\}$$

The first component, $1 - e^{-ax}$, is common to both the leader and laggard; it represents the baseline probability of success. The second component, $\max\{0, 1 - e^{\eta(n - \hat{n})}\}$, is specific to the laggard; it represents the additional probability of success to a laggard that benefits from spillovers.

Given $P(\cdot)$, Hashmi solves the Bellman equation. He then plots $x(\cdot)$ against $\alpha$ under different values of $n$. The results indicate that the relationship between $x$ and $\alpha$ (i.e., the relationship between innovation and competition) depends on $n$. Specifically, for small values of $n$, the relationship is approximately monotonically increasing. For intermediate values of $n$, the relationship is an inverted-U. And when $n$ is large, the relationship is monotonically decreasing.

1.4.2 Empirical models

Early empirical studies focused on scale—firm size and market concentration—and its effect on innovation. New empirical research is expressly concerned with the effect of
In addition to a different, albeit somewhat related, policy question, the new literature differentiates itself via its more sophisticated econometric techniques, its measures of competition and innovation and its quantity and quality of data.

Broadly speaking, two classes of model exist in the literature: linear and non-linear. That is, the literature may be divided in terms of how the relationship between competition and innovation is specified. Nickell (1996) and Blundell et al. (1999) estimate a linear specification. ABBGH, Correa (2012), Correa and Ornaghi (2014) and Hashmi (2013) estimate non-linear specifications.

**Review of the literature**

Both Nickell (1996) and Blundell et al. (1999) estimate linear models and find a positive relationship based on a panel of publicly traded firms in the United Kingdom. Nickell’s principal measure of competition is a firm’s average operating margin, whereas Blundell, Griffith and Van Reenen measure competition in terms of individual market share, market concentration and import penetration. For innovation, Nickell uses productivity as a proxy, and Blundell, Griffith and Van Reenen use survey-based innovation counts.

The non-linear class of models, all based on ABBGH, generate a range of qualitative results. ABBGH find an inverse-U shape; Correa (2012), who uses the same data and baseline model as ABBGH, finds a monotonically increasing relationship for some years of the sample, but no relationship otherwise; Correa and Ornaghi (2014) finds a monotonically increasing relationship; and Hashmi (2013) finds a monotonically negative relationship.

---

8One might argue that these strands of literature are one in the same, particularly if competition is dampened by scale. However, changes in size alone are not enough impact competition.
tive relationship. The difference in results for the non-linear class of models in striking given the strong similarity in data and methods. This bears further discussion.

As mentioned above, empirical investigations of a non-linear relationship between competition and innovation are derivatives of ABBGH. ABBGH start with the following conditional mean function:

$$E[p|c] = e^{g(c)},$$

where $p$ is the number of citation-weighted patents, $c$ is competition and $g(\cdot)$ is some function to be estimated. It is assumed that $p$ follows a Poisson process. Below is scatter plot of their data with an exponential quadratic overlay.\(^9\) ABBGH estimate their model with 354 industry-year observations. Data are on publicly traded manufacturing firms in the United Kingdom over the period 1973 to 1994.

\(^9\)This graph is taken from Aghion et al. (2005).
Correa (2012) investigates the results of ABBGH further using the same sample of data and empirical formulation. The difference is that he allows for a structural break in the data. He argues that this is appropriate because the establishment of the United States Court of Appeals for the Federal Circuit (henceforth CAFC) in 1982 made it effectively easier to have a patent granted.

Correa takes two approaches to test the structural break hypothesis. The first approach is a Chow test. The base model estimated is the same as in ABBGH:

\[
p_{jt} = \exp \left\{ \beta_0 + \beta_1 c_{jt} + \beta_2 c_{jt}^2 + \phi \hat{v}_{jt} + \delta_1 D_{\tau} c_{jt} + \delta_2 D_{\tau} c_{jt}^2 + \sum_{j=1}^{17} \alpha_j D_j + \sum_{t=1973}^{1994} \gamma_t D_t + u_{jt} \right\}
\]

where \( D_{\tau} = 1 \) for all \( t \geq \pi \), 0 otherwise; \( c_{jt} \) is the level of competition for industry \( j \) at time \( t \), measured as one minus the industry average price-cost margin; \( \hat{v}_{jt} \) is the residual for industry \( j \) at time \( t \) from regressing the competition index on policy and foreign-industry instruments (i.e., endogeneity is accounted for with a control function approach); and last two terms are industry- and time-fixed effects. The Chow test corresponding to the above is

\[
H_0 : \; \delta_1 = \delta_2 = 0
\]

\[
H_1 : \; \text{otherwise}
\]

The null hypothesis of time stability at \( t = 1983 \) is rejected at the 5% level of significance. Below is a graphic that shows the change in the relationship between competition and innovation before and after the CAFC reform.\(^\text{10}\)

\(^{10}\)The plots below are taken from Correa (2012)
The second approach is a Sup-Wald test, which is a statistical “brute-force” method to find structural breaks in the data. The hypothesis test is given by

\[ H_0 : \delta_1 = \delta_2 = 0 \]

\[ H_{1T} : \delta_{jt} = \begin{cases} \delta_{j1}(\pi) & \forall t = 1, \ldots, T\pi \\ \delta_{j2}(\pi) & \forall t = T\pi + 1, \ldots \end{cases} \]

for constants \( \delta_{j1} \) and \( \delta_{j2} \) and break point \( \pi \in (0, 1) \), where

\[
\sup_{\pi \in \Pi} W_T(\pi) = \arg\max \left[ W(\pi_{\Pi}), W(\pi_{\Pi}) + 1, \ldots, W(\pi_{\Pi}) \right]
\]

for \( \Pi = [0.05, 0.05] \). With this approach, Correa finds only one structural break at year 1981. Correa gives several reasons why this year, instead of 1982, was detected. Perhaps his most convincing argument is that the political discussion to establish the
CAFC began in 1979, which in turn may have altered patent behavior before the CAFC was officially established.

The indication of a structural break in the data by both tests leads Correa to test the joint hypothesis

\[ H_0 : \beta_1 = \beta_2 = 0 \]

\[ H_1 : \text{otherwise} \]

before and after the determined structural breaks of 1981 and 1982. In both cases, the null hypothesis is rejected (fails to be rejected) at the 5% significance level before (after) the structural break. In other words, before (after) the establishment of CAFC, the relationship is statistically significant (not statistically significant). Finally, in both cases he finds that the relationship between competition and innovation is monotonically increasing before the structural break.

Hashmi (2013) and Correa and Ornaghi (2014) also revisit ABBGH. Unlike Correa (2012), however, slight modifications are to the model and different data are used. These papers have the following in common: both use (1) a negative binomial instead of a Poisson specification for the mean function; (2) data on publicly traded manufacturing firms in the United States and; (3) the Lerner index as a proxy for competition. The reason for (1) is data driven; both find overdispersion in their data (i.e., the mean of patent counts is different than its variance).

Despite the similarities, the qualitative results vary. Correa and Ornaghi (2014) find a monotonically positive relationship using total factor productivity growth, labor productivity and citation-weighted patent counts as a proxy for innovation, while Hashmi (2013)—using only citation-weighted patent counts as a proxy for innovation—finds a monotonically “mild,” but negative relationship.
1.5 Related Empirical Literature Outside IO

1.5.1 Using natural experiments to identify the effect of competition

The empirical IO literature on competition and innovation has exclusively focused on concentration and profitability to infer the level of competitiveness of an industry, and largely patents and R&D to capture the level of innovation. Another literature, however, investigates the link between “discrete” changes in the competitive environment and productivity, where a “discrete” change means the competitive environment underwent a clear, identifiable transformation. Closely following the survey of Holmes and Schmitz (2010), this section presents a brief review of the literature.

Review of the literature

Before getting into the literature, I would first like to draw attention to some comments made by Holmes and Schmitz regarding the measurement of competition and how competition affects productivity. Holmes and Schmitz contend that concentration and profitability are inadequate at identifying structural changes in the competitive environment. In fact, they conjecture that these measures have the potential to mislead, to which they give the following example. Suppose an industry is initially composed of small, relatively unproductive firms where there is a strong barrier to entry, but then the barrier is lifted. In particular, imagine the government lifts a trade barrier and subsequently a large, highly productive firm enters. From the perspective of a researcher who observes only market shares and profitability, she would then conclude that the industry became less competitive due to a substantial increase in concentration and profitability. This, however, is counter to the notion that less barriers stimulate competition. They further this argument by pointing out the selection
effect competition, which is that relatively unproductive firms are prone to exit in the face of greater competition. This has two implications. One, market concentration can *increase* in a more competitive state. Two, if productivity is positively correlated with profitability, then average profitability will also *increase* in a more competitive state.

That said, how does competition affect productivity? Holmes and Schmitz claim that no workhorse model exists that can explain why or how competitive pressure induces firms to be more productive. The body of evidence strongly suggests that it does, however. For example, Matsa (2011) found that incumbent retailers made substantial improvements to their inventory control systems once Wal-Mart entered their market, and that the increase in productivity cannot be attributed to an increase in market or average firm size\(^1\)—demand did not all of a sudden increase for the existing firms, nor did the existing firms substantially increase in scale. The observed gain in productivity was therefore almost surely due to an increase in competitive pressure.

Opportunity cost is another way in which competition can bolster productivity. For example, Schmitz (2005) found that plant managers were reluctant to adopt new managerial practices from fear of lost profits via a strike. From this, he argues that the competitive process—which in the absence of innovation tends to shrink margins over time—will reduce the opportunity cost of lost profits and thereby spur investment into new, more efficient forms of management.

Major shifts in the competitive landscape are perhaps the best way to identify any effect of competition, if at all, on innovation. Consider first the example of

\(^{11}\)Theoretically, the incentive to lower costs increases with firm scale. Intuitively, the larger the firm is, the more it can save by reducing costs.
Holmes and Schmitz (2001). They examine the effect of railroad transportation on water shipping in the 19th and 20th centuries. Before the advent of economically feasible transportation by railroad in the US (1850s), freight transportation by water was effectively the only way to ship cargo across the nation. This meant that ports, which were separated by great distances, not only had tremendous market power, but also had the incentive to keep it. Their market power was heavily weakened, however, when railroads became a viable alternative for transportation.

Railroads weakened the market power of ports in two ways. On one hand, railroads gave consumers easier accessibility to other ports. So, if a consumer was not pleased with the price or service of a port, it could, with relatively little expense, use a train to ship its cargo to another port. On the other hand, railroads themselves could in some cases bypass the services of water transportation altogether. The threat that railroads presented to ports manifested itself in an effort by the latter to increase productivity. In particular, longshoremen relaxed what were very restrictive labor practices that led to long docking times for ships. These restrictive practices included a rule that forbade the use of machinery to load or unload ships and a rule that forbade direct transfer of cargo from boat to ship and ship to boat.

In two other papers where the competitive environment significantly changed, Galdon-Sanchez and Schmitz (1976) and Schmitz (2005) examine the effect of Brazil’s entry into the lower Great Lakes iron ore market during the 1980s. They argue that, before Brazil’s entry, there was a tendency for monopoly since iron ore markets were characterized by few and distant producing locations and high transportation costs. Market power was evident since several organizations, such as unions and government at the state, county, township and school board levels, exercised their power to extract as much of the surplus from iron ore producers as possible.
The market power of iron ore producers around the Great Lakes would not last, however. Due to a substantial decrease in transportation costs (and shrinking markets elsewhere for Brazil), Brazil entered the iron market around the Great Lakes in the 1980s. This put tremendous price pressure on the domestic iron ore producers and, in turn, pressure to improve labor productivity. Labor productivity actually doubled in the mid 1980s, and Galdon-Sanchez and Schmitz show that the source of productivity growth was not due to the closing of inefficient mines or increases in scale, but rather surviving mines that made investments to lower costs (changes to shift schedules and repair job classifications were made to reduce downtime).

Finally, Syverson [2004] investigates the effect of spatially dense competition on the distribution of productivity in the United States ready-mix concrete industry. He finds that more densely clustered markets exhibit higher average productivity and lower productivity dispersion. The fundamental reason for this, he argues, is that more densely clustered markets lower switching costs for consumers. As such, an inefficient firm is more likely to exit an industry that is highly dense, meaning average productivity and productivity dispersion will increase and decrease, respectively, in more competitive markets.

1.6 Problem Statement

In light of IO theory and evidence, it is clear that no definitive statement can be made about the competition-innovation relationship.

“Discrete” theoretical models of competition—Arrow (1962), Gilbert and Newbery (1982) and Reinganum (1983)—disagree on whether firms with market power are more or less likely to innovate than firms without; both Arrow and Reinganum find that market power acts as a disincentive to innovate, while Gilbert and Newbery argue
that the threat of foregone rents is enough to keep innovation alive in markets with substantial market power.

“New” theory—where the measure of competition ranges from the expected date of innovation by a rival, the number of firms in a symmetric industry and the degree of product substitutability in a market with differentiated goods—paints a complex picture of competition and innovation. Kamien and Schwartz (1976) find that the optimal introduction date of innovation first declines with more intense rivalry, but then increases due to convexity in the development cost function. Loury (1976) finds that when the number of firms in a symmetric industry increases, individual R&D investment goes down, but the introduction date of innovation accelerates. AHHV find that greater product substitutability leads in most cases to greater aggregate R&D intensity, but if the step-size of innovation is large, too much can harm it. ABBGH find that, regardless of the size of innovation, product substitutability and aggregate R&D intensity share an inverted-U relationship. And Hashmi (2013) finds that as the industry technological gap varies, so too does the relationship between product substitutability and aggregate R&D intensity: a small gap corresponds to a positive relationship; an intermediate gap corresponds to an inverted-U; and a large gap corresponds to a negative relationship.

The empirical evidence is equally mixed. Early empirical studies, while not concerned with competition per se, find little evidence of any statistical relationship between firm size/market concentration and R&D activity. In contrast, the new literature finds evidence of market power, as measured by an industry’s average operating margin, both thwarting and facilitating innovation. The range in qualitative results for non-linear specifications—ABBGH, Correa (2012), Correa and Ornaghi (2014) and Hashmi (2013)—is particularly striking given the similarity in data and methods.
Moving outside the literature of theoretical and empirical IO, we find strong evidence of competition driving innovation. Schmitz (2005) shows that iron ore producers in the Great Lakes region nearly doubled their labor productivity in the 1980s after Brazil entered the market with substantially lower prices. Holmes and Schmitz (2001) find evidence that water shipping companies improved their labor productivity once railroad transportation became a viable alternative in the late 19th and early 20th centuries. And Matsa (2011) shows that incumbent retailers made significant investments to improve their inventory control systems after Wal-Mart entered their market.

The trait shared among the literature cited immediately above is how an increase in competition is captured. It does not capture “more competitive” in terms of how many firms there are, profitability or concentration. Instead, it captures “more competitive” in terms of the threat rivals pose to one another. This fundamental aspect of the competitive process is what makes competition effective, and yet it fails to be captured directly in the IO literature. Take the following excerpt from Aghion et al. (2001), for example:

The [product] substitutability parameter $\alpha$ is our measure of the degree of product market competition in each industry. Although $\alpha$ is ostensibly a taste parameter, we think of it as proxying the absence of institutional, legal and regulatory impediments to entering directly into a rival firm’s market by offering a similar product. Under this interpretation $\alpha$ reflects in particular the influence of anti-trust policy.

They then go on to say:

In our model $\alpha$ corresponds to standard measures of competition. For
example, it is a monotonically increasing transformation of the elasticity of substitution in demand \( (1/(1 - \alpha)) \) between the two rivals’ outputs in any industry. Given a firm’s share \( \lambda \) of industry revenue, \( \alpha \) is also a monotonically increasing transformation of the elasticity of demand \( (1 - \alpha \lambda)/(1 - \alpha) \) faced by the firm. Furthermore, given \( \lambda \), \( \alpha \) is a monotonically decreasing function of the measure of market power used in the related empirical research by Nickell (1996), namely the share of profits in value added, which in this model is \( (1 - \alpha)/(1 - \alpha \lambda) \).

This excerpt was chosen because it represents what is common in the IO literature, which is a conceptually vague definition of competition. Like the example above, the measurement choice is often justified on the grounds that it is standard or that it has a monotonic relationship with market power. But why are these measures standard? And what is market power? I address these questions in Chapter 2.

Standard theoretical measures of competition are, in fact, aspects of the competitive environment, not measures of “more competitive.” Schmitz (2005) provides a clear illustration. Recall that when Brazil entered the iron ore market near the Great Lakes region, domestic labor productivity nearly doubled. What drove this? Clearly, there was no change to the degree of product substitutability—iron ore is iron ore—so this is ruled out. What about the number of firms? Assuming no manufacturers simultaneously exited, the number of firms must have increased. But an increase in the number of firms is unlikely to have been the reason why productivity doubled. For example, suppose instead some other manufacturers entered the market but they offered the prevailing price. Then there would have been little to no pressure for domestic manufacturers to raise their productivity. Thus, in order for competition to effectively increase, the threat rivals pose to one another must also increase.
Like their theoretical counterparts, standard empirical measures of competition fail to capture “more competitive.” Consider Syverson (2004). Syverson finds that when rivals are more densely clustered in space—meaning consumer transportation costs are lower—the distribution of productivity tends to be truncated from the bottom, i.e., the least efficient firms exit. In turn, average productivity increases. The implication of this result is twofold. First, to the extent that productivity is correlated with the Lerner index, a more competitive environment would indicate higher margins. Second, the decrease in the number of firms due to exit, all else equal, will increase market concentration. In other words, this example illustrates a case where the Lerner and Herfindahl-Hirschman indices—the most common empirical measures of competition in the literature—would fail to identify an increase in competition.

I seek to address the issue just described. Namely, I develop a framework where the definition of competition is precisely defined and in terms of the threat rivals directly pose to one another. I specifically follow Shapiro (2012) and define a “more competitive” industry as one where a firm stands to lose greater sales to its rivals should it offer inferior value to consumers. This definition is then applied directly in my theoretical and empirical models.

In addition to the above, I also address the empirical measurement of innovation. New empirical IO models of competition and innovation have gravitated toward patents as a proxy for innovation. I argue that patent statistics are inadequate based on the nature of patents, firm behavior and evidence that they are non-monotonically related to value. As an alternative, I measure innovation in terms of productivity, which is consistent with long-run economic growth. Chapter 2 now follows.

12See Hall (1990) for a decomposition
Chapter 2

Competition and Innovation:
A Theoretical Framework

2.1 Introduction

Antitrust authorities focus on three types of cases (Shapiro (2005)): price fixing, mergers and acquisitions and monopolization. In terms of competition policy, the competition-innovation literature is most closely related to the latter two vis-à-vis its attention to market power. Market power in the textbook sense, however, is substantially different from the antitrust view of market power in the United States. The former is technical and quantifies market power in terms of market shares and profit margins, while the latter defines enhanced market power as a reduction in consumer welfare stemming from “diminished competitive constraints or incentives.”

I adopt a measure of competition consistent with the antitrust view of market power and examine its effect on innovation.

The root of enhanced market power per the Department of Justice and Federal Trade Commission is consistent with a decrease in rival discipline, or the threat rivals pose to one another. I capture this using the definition of “more competitive” advanced by Shapiro (2012): an industry is more competitive if a firm stands to lose greater sales to its rivals should it offer inferior value to consumers. The emphasized parts of this definition highlight two key aspects. One, sales “switch hands” based

---

1Department of Justice Horizontal Merger Guidelines as of August 19, 2010.
on the value rivals offer to consumers. Most of the models I develop directly capture this by imposing a zero-sum property on competition. While this assumption may seem restrictive, it in fact allows us to separate the disciplinary effect of competition on innovation from other confounding factors. That is, if a firm innovates and subsequently sees an increase in its sales, the zero-sum property allows us to identify what part of the increase is due to a “business stealing” effect and what part of it is due to, perhaps, an increase in the size of the market. Two, a firm can only steal sales if it offers relatively greater value to consumers. This is explicit in each of my models. For example, in a game with stochastic innovation, I make the assumption that a firm can only steal profit from its rival if it is the sole innovator.

The rest of Chapter 2 proceeds as follows. Section 2.2 discusses the concept of market power from an antitrust and academic perspective. Section 2.3 gives an overview of traditional measures of competition. Section 2.4 discusses innovation in theory and in practice. Section 2.5 examines game theoretic models of competition and innovation where greater competition corresponds to a more contestable market. And section 2.6 develops the basic empirical strategy to test my theoretical prediction of a monotonically positive relationship between competition and innovation.

### 2.2 Market Power

Market power as a concept is central to antitrust law and the theoretically grounded structure-conduct-performance paradigm of industrial organization. There are stark differences, however, between how this concept is currently applied in practice and how it is applied in theory.

From a theoretical perspective, market power is the extent to which a firm can price above marginal cost—the larger the gap is between a firm’s price and its marginal
cost, the more market power that firm has. Antitrust law in the United States instead defines enhanced market power as a reduction in consumer welfare stemming from “diminished competitive constraints or incentives.” More specifically, the 2010 Horizontal Merger Guidelines set forth by the United States Department of Justice (DoJ) and Federal Trade Commission (FTC) states that a “merger enhances market power if it is likely to encourage one or more firms to raise price, reduce output, diminish innovation, or otherwise harm consumers as a result of diminished competitive constraints or incentives.”

Thus, while the theoretical definition of market power focuses strictly on price competition, the antitrust definition focuses on rivalry. This was not always the case, however. From 1982 to 2010, the Horizontal Merger Guidelines defined market power as the “ability of one or more firms [to profitably] maintain prices above competitive levels for a significant period of time”—a reflection of its theoretical counterpart.

The theoretical definition of market power and its historical use by antitrust authorities in the United States is attributable to the then dominant structure-conduct-performance paradigm. The structure-conduct-performance framework posits a direct relationship between market structure, firm conduct and performance. For example, concentrated industries (market structure) foster collusion (firm conduct) which in turn leads to less output (performance). The connection to antitrust is thus clear regarding its power to affect market structure. But where does market power fit in? It turns out that there is a theoretically direct correspondence between market power, market structure and social welfare.

The link between market power, market structure and welfare was first demonstrated by Lerner (1934). In his 1934 publication in the The Review of Economic Stud-

\[\text{See HMG (2010) for the 2010 Horizontal Merger Guidelines.}\]
“The Concept of Monopoly and the Measurement of Monopoly Power,” Lerner observed that when a single, profit maximizing firm chooses its output optimally, it generates an allocative inefficiency that is inversely related to its price-elasticity of demand. Specifically, he found that the less elastic demand is (in magnitude), the greater is the gap between the output that is produced and the output that is socially desirable. Showing this requires only weak assumptions, namely a downward-sloping demand curve and an upward-sloping marginal cost curve.

Formally, let \( P(Q) \) and \( C(Q) \) denote a single firm’s demand and cost curve, respectively. Then the profit function, \( \pi(Q) \), is given by

\[
\pi(Q) \equiv P(Q)Q - C(Q).
\]

The profit maximizing level of output, \( Q_m \), must satisfy

\[
L(Q_m) \equiv \frac{P(Q_m) - C'(Q_m)}{P'(Q_m)} = \frac{1}{|\epsilon(Q_m)|},
\]  

(2.1)

where \( L(Q) \) is the Lerner index, \( C'(Q) \) is the firm’s marginal cost curve and \( \epsilon(Q) \equiv \frac{1}{P'(Q)} \frac{P(Q)}{Q} \) is the firm’s price-elasticity of demand. From (2.1) it is clear that a decrease in the price-elasticity of demand must be met with an increase in the wedge between price and marginal cost. Moreover, since this wedge increases with a reduction in output (given the assumptions on \( P(\cdot) \) and \( C'(\cdot) \)), the gap between the output that is produced, \( Q_m \), and the output that is socially desirable (i.e., the level of output \( Q_s > Q_m \) where \( P(Q_s) = C'(Q_s) \)) must also increase.

Taken at face value, the divergence between price and marginal cost may be viewed as a divergence from a hypothetical perfectly competitive equilibrium. Lerner observed this and consequently took a larger wedge to mean greater “monopoly power.”

---

3Note, the marginal cost curve could be constant and Lerner’s result would still attain.
The resilience of the Lerner index as a measure of market power is not surprising; as a matter of application, it has several appealing features. One, the Lerner index resonates with the intuition that more market power should increase prices, which predicts. Two, if it can be reasonably assumed that marginal cost is constant, then calculation of the Lerner index requires data only on average costs, which are observable. Three, “deep” parameters of competition, like the degree of product substitutability, can typically be shown to have a monotonic relationship with the Lerner index. And four, the Lerner index provides rigor to the otherwise subjective concept of market power.

The conveniences of the Lerner index are outweighed by its theoretical faults, however. For example, the Lerner index ignores cost-minimizing behavior; is derived within a static framework and thus cannot account for dynamic factors like technological change and learning by doing; ignores non-price competition and; neglects the idea that departures from marginal cost pricing may be attributable to economies of scale and/or the need to cover fixed costs. Each of these scenarios could result in a

---

4To be clear, the source of market power is the inverse of the price-elasticity of demand.
5There is, for example, an entire text dedicated to estimating market power. See “Estimating Market Power and Strategies” by Jeffrey M. Perloff, Larry S. Karp and Amos Golan, 2007.
6Note that market power need not be defined in terms of the price-elasticity of demand. More generally, market power (as an expression) is a construct that evokes some sense of monopoly. And because monopoly is never compatible with consumer welfare, anything that more closely resembles monopoly should be worse for consumers (be it in terms of price, quality, customer service, etc.).
7Multiplying both the numerator and denominator by $Q$, the Lerner index becomes just a ratio of profits to sales. There is a caveat, however. We really need economic, not accounting, profit.
8See Elzinga and Mills (2011) for a discussion of the Lerner index and its faults as a measure of market power.
greater wedge between price and marginal cost, and yet none of them are indicative of meaningful market power.

Consider, for example, the effect of cost-minimizing behavior, which may be industry driven, the result of product maturity or the result of increased competitive pressure (e.g., an import tariff is lifted). Let \( c(t) \) and \( p(c(t), t) \) denote some firm’s period \( t \) marginal cost and price, respectively, and suppose that (i) price is the only strategic variable, (ii) \( \frac{dc}{dt} < 0 \) is observed for some interval of time \( t \in [t_0, t_1] \) and (iii) \( \frac{\partial p}{\partial c} \geq 0 \). The firm’s Lerner index is given by

\[
L(t) = 1 - \frac{c(t)}{p(c(t), t)} \geq 0.
\]

Differentiating with respect to \( t \), we find

\[
\dot{L} = (1 - L) \left( \frac{\dot{p}}{p} - \frac{\dot{c}}{c} \right) = (1 - L) \left( \frac{\frac{\partial p}{\partial c} \dot{c} + \frac{\partial p}{\partial t} p}{p} - \frac{\dot{c}}{c} \right)
\]

where, in general, \( \dot{x} \equiv \frac{dx}{dt} \) denotes the time derivative of \( x \). For simplicity of exposition, assume that there are no price-related macroeconomic shocks over the relevant interval, \( i.e., \frac{\partial p}{\partial t} = 0 \). Then the change in the Lerner index over the time interval \([t_0, t_1]\) will be negative if and only if the elasticity of price with respect to marginal cost is elastic, \( i.e., \right \frac{\partial p}{\partial c} \frac{c}{p} > 1 \).

To put this into some context, suppose the product space that the firm competes in is mature; that is, prices are stable and higher margins are achieved primarily by lowering costs. In this scenario, it is not unreasonable to expect price reductions to be relatively minimal. Thus, we would observe an increase in the Lerner index over
the relevant interval of time unrelated to any significant change in the competitive environment.

Suppose now that there is a change in the competitive environment. For example, an import tariff is lifted which forces domestic firms to compete with its now more efficient rivals. In order for the domestic firms to compete effectively, they will have to lower their costs. Two possibilities may arise from this. One, if the subsequent reduction in prices that arise from lowered costs is inelastic on average, we will see an increase in the average Lerner index. Two, a subset of inefficient firms may have to exit the industry, which will show up as an increase in average productivity. In turn, since productivity is positively correlated with the Lerner index, we will also see an increase in the average Lerner index.

The theoretical arguments for why the Lerner index might fail to detect market power are well founded. Suppose, though, that the scenarios just outlined occur only a fraction of the time. Even if this is the case, changes in the Lerner index over small windows of time are unlikely to be informative. This much is acknowledged by the 1982-2010 Horizontal Merger Guildelines: “the ability of one or more firms to [profitably maintain] prices above competitive levels for a significant period of time is termed ‘market power.’” But even this criterion might not be enough. Hay (1992) argues convincingly that for the Lerner index to be a useful measure of market power, market power must not only be non-transitory but also significant, i.e., the departure from price and marginal cost must also be sufficiently large. Indeed, a firm that prices above marginal cost yet still earns a negative economic profit (i.e., price is not high enough to cover fixed costs) is unlikely to pose any threat to consumer welfare.

\[\text{See Syverson (2004) for an example of this.}\]
\[\text{See Hall (1988) for the relationship between productivity and market power.}\]
Given the theoretical limitations of the Lerner index as an indicator of meaningful market power, I adopt the following definition of “more competitive:” an industry is “more competitive” if a firm stands to lose greater sales to its rivals should it offer inferior value to consumers. This definition was proposed by Shapiro (2012) with the view that “more competitive” should be reserved for “market characteristics that correspond to greater rivalry to serve the needs of customers.” It both resonates with the intuition that more contestable markets spur rivalry—which he calls the “contestability principle”—and is in fact consistent with the current practice of antitrust law. In applying this definition, I find a robust positive relationship between competition and innovation.

The next section gives a brief overview of other measures of competition used in the literature.

### 2.3 Traditional Measures of Competition

In the previous section, I gave a brief overview of the history of market power as a concept and its relationship to the Lerner index. In this section, I discuss other measures of competition commonly seen in the literature. These measures include the Herfindahl-Hirschman index, number of firms in an industry and product substitutability.

#### 2.3.1 Market Concentration

One of the most commonly used empirical measures of competition is the Herfindahl-Hirschman index (HHI). It measures the degree of concentration within an industry
by summing each firm’s squared market share, i.e.,

\[ HHI = \sum_{i} s_{i}^{2}, \]

where \( s_{i} \) is firm \( i \)'s market share. The appeal of this index is that it not only gives more weight to larger firms, it also increases in fewness of firms. The latter is easily seen by its decomposition:

\[
HHI = \frac{n}{Q^{2}} \left( \frac{1}{n} \sum_{i} q_{i}^{2} \right), \quad Q \equiv \sum_{i} q_{i}
\]

\[
= \frac{n}{Q^{2}} \left( s^{2} + \bar{q}^{2} \right), \quad s^{2} \equiv \frac{1}{n} \sum_{i} (q_{i} - \bar{q})^{2} \quad \text{and} \quad \bar{q} \equiv \frac{1}{n} \sum_{i} q_{i}
\]

\[
= \frac{n\bar{q}^{2}}{Q^{2}} \left( 1 + \frac{s^{2}}{\bar{q}^{2}} \right)
\]

\[
= \frac{1 + \nu^{2}}{n}, \quad \nu \equiv \sqrt{\frac{s^{2}}{\bar{q}^{2}}}
\]

where \( \nu \) is the sample coefficient of variation (i.e., the inverse signal-to-noise ratio) and \( n \) is the number of firms. Its popularity is therefore not surprising—it accords with the conventional wisdom that industries with larger and fewer firms are less competitive.

It can in fact be shown in a general Cournot setting with \( n \) cost-asymmetric firms and homogeneous goods that there is a direct relationship between the HHI and the Lerner index. Let \( q_{i} \) and \( C_{i}(q_{i}) \) denote firm \( i \)'s output and cost function, respectively; \( Q \equiv \sum_{i} q_{i} \) total output; and \( P(Q) \) the inverse market demand curve. The profit function facing firm \( i \), \( \pi_{i}(Q) \), is given by

\[
\pi_{i}(Q) = (P(Q) - C(q_{i}))q_{i}.
\]

\[11\text{See }\text{Hirschman (1945)} \text{ for a derivation.}\]
The first-order condition implies

\[
\frac{P(Q^*) - C_i'(q_i^*)}{P(Q^*)} = \frac{P'(Q^*)}{P(Q^*)} q_i^* \\
\iff L_i^* = \frac{s_i^*}{|\epsilon^*|} \\
\iff s_i^* L_i^* = \frac{(s_i^*)^2}{|\epsilon^*|} \\
\iff \sum_i s_i^* L_i^* = \frac{\sum_i (s_i^*)^2}{|\epsilon^*|} \\
\iff \sum_i s_i^* L_i^* = \frac{HHI^*}{|\epsilon^*|},
\]

where, in equilibrium, \(L_i^* \equiv \frac{P(Q^*) - C_i'(q_i^*)}{P(Q^*)}\) and \(s_i^* \equiv q_i^*/Q^*\) are firm \(i\)'s Lerner index and market share, respectively; \(\epsilon^* \equiv \frac{1}{P'(Q^*)} \frac{P(Q^*)}{Q^*}\) is the market elasticity of demand; and \(HHI^* \equiv \sum_i (s_i^*)^2\) is the Herfindahl-Hirschman index.

As can be seen, the share-weighted average Lerner index shares a direct relationship with the HHI, and the firm-specific Lerner index shares a direct relationship with market share. In turn, market share and market share concentration are a manifestation of market power. But such a notion is pass; an increase in concentration or market share may signal nothing more than a firm or set of firms performing at a relatively high level (Demsetz (1973)).

The HHI also neglects to recognize that market power is relative; i.e., the market power of a firm depends on its as well its rivals’ competitive positions. To illustrate, suppose that the competitive position of a firm is determined only by its market share. Now consider two scenarios: (a) four firms with market shares 2/3, 1/9, 1/9 and 1/9 and (b) two firms with equal market shares. A simple calculation shows that the HHI equals 13/27 < 1/2 in scenario (a) and 1/2 in scenario (b). Application of the HHI would therefore indicate that scenario (a) is more competitive. But is it? While there are less firms in scenario (b), scenario (a) is highly unbalanced compared to
scenario (b). In particular, one firm has the vast majority of market share in scenario (a), while two firms are competitive equals in scenario (b). Both scenarios of course exhibit high concentration, but it could argued that scenario (a) is theoretically more harmful to consumer welfare because it is much closer to a textbook monopoly.

Finally, the HHI is difficult to apply in practice. Recall that the relationship between the Lerner index and the HHI is derived under the assumption that goods are homogeneous. Now suppose we have data at the firm level, but not at the line-of-business level (as if often the case). By comparing firms in terms of their sales, we implicitly assume that the firms—be they single- or multi-product firms—sell identical product(s). This is unlikely; even firms in the same industry have substantial differences in terms of the variety and quality of products sold. Thus, even if the HHI was a theoretically reasonable measure of competition, the lack of granular data will undermine its effectiveness.

2.3.2 Number of Firms and Barriers to Entry

The number of firms in an industry and barriers to entry are directly related; in a model with endogenous entry, higher barriers to entry amounts to less firms entering the market. I consequently restrict attention to the case of a fixed market structure.

The number of firms in an industry is a common way to think about how competitive an industry. Its use is not based on some monotonic relationship with market power, however. Rather, it is used because it directly measures how many comp-

\[ L^*_i = L^* = 1 - \frac{C'(q^*(n))}{P(nq^*(n))}, \]

where \( q^* \) is the equilibrium quantity for all \( i \). Differentiating this with respect to \( n \) (and simplifying
petitors there are in a market. More competition, however, is not equivalent to more competitive. For example, if a firm exits its industry because it was unable to compete effectively, this does not imply that the industry is now more competitive. Instead, it just reflects the fact that the competitive process selects the highest performing firms.

Application of this measure also fundamentally depends on the relevant market. If the relevant market is not approximated well, then the number of firms will be a noisy measure of more competition. Thus, both as a measure in theory and in practice, the number of firms in an industry cannot reasonably identify an industry’s level of competitiveness.

2.3.3 Product Substitutability

When inter-industry competition is the focus and products are differentiated, product substitutability is perhaps the most common theoretical measure of competition. It is easy to understand why if price is the strategic variable: closer substitutes imply that a firm’s demand is more sensitive not only to its own price, but to the prices of its rivals as well. Greater product substitutability will consequently put greater pressure on firms to lower their prices.

Another “feature” of product substitutability that drives its use is its relationship with market structure. Specifically, parameterizations of product substitutability are...
always such that, in the upper limit, each firm’s demand function becomes infinitely price-elastic and, at zero, solely a function of one’s own price. That is to say, in the upper limit, the market resembles a perfectly competitive industry, and at zero it resembles an industry of local monopolists. In fact, Vives (2008) shows under various demand functions that when competition is Bertrand, products are differentiated and entry is restricted, the Lerner index is monotonically decreasing in product substitutability, supporting the notion that product substitutability is a valid measure of market power.

From a classical view of price competition, product substitutability is a sensible measure of competitive pressure. However, it does not embody the principle of contestability advanced by Shapiro. For example, two competing products can be highly substitutable (e.g., similar in terms of features, quality, service, functionality, etc.), but switching costs (real or perceived) may be so high that neither firm presents a real threat to the other. This could be due to brand loyalty, marketing, reputation, location, the time it takes to learn a new product, etc. Whatever the reason, if both firms act on the belief that the neither firm presents a significant threat, high substitutability will not matter as far as effective competition is concerned.

Of course, in a theoretical world of price competition, one might argue that an increase in product substitutability is tantamount to an increase in contestability. In particular, one might argue that when price is the only strategic variable, a higher degree of substitutability implies that lower-price firms will capture even more sales from higher-price ones, thereby making the market more contestable. This argument, however, misses the point. Contestability relates to the ability of rivals to contest one

\[\text{\textsuperscript{13}}\] However, when entry is allowed, Vives shows that the monotonic relationship between the Lerner index and product substitutability breaks down.
another’s sales *despite* how similar products and/or services are. The distinction is subtle, but it matters. In the realm of mergers and acquisitions, an antitrust authority will permit a merger if the merger is expected to increase competitive incentives, not because it will serve to make the industry more homogeneous.

In the next section, I present a standard model of Bertrand competition with differentiated products, restricted entry and process innovation. This model illustrates a non-robust relationship between R&D intensity and product substitutability.

### The Relationship Between Product Substitutability and R&D Effort: An Example

The following model is a generalization of [Singh and Vives (1984)] and [Qiu (1997)].

Consider a non-cooperative, two-stage game of price competition and process innovation. There are two firms. In the first stage, firms simultaneously choose their levels of R&D, $x_i$, $i = 1, 2$, to reduce their constant marginal cost of production, $c$. In the second stage, firms simultaneously choose their prices, $p_i$, $i = 1, 2$ to maximize gross profit.

The representative consumer’s preferences are characterized by the following utility function

$$u(q_1, q_2) = \alpha(q_1 + q_2) - \frac{1}{2} \left(q_1^2 + q_2^2 + 2\beta q_1 q_2\right),$$

where $q_i$ is firm $i$’s production level, $\alpha > 0$ and $\beta \in (0, 1)$. The parameter $\beta$ measures the degree of product substitutability. The consumer maximizes its utility subject to its income constraint, $\sum_{i=1}^{n} p_i q_i < I$, where $I$ is income. The first-order conditions are

$$p_i = \alpha - q_i + \beta \sum_{j \neq i} q_j, \quad i = 1, 2, j \neq i$$
Solving this system for quantities, we get the following demand functions:

\[ q_i(p_i, p_j) = \frac{\alpha}{1 + \beta} - \frac{p_i - \beta p_j}{1 - \beta^2}, \quad i = 1, 2, j \neq i. \]

Before moving on to the firm’s profit maximization problem, notice a few things about \(\beta\). First, when prices are identical and \(\beta\) approaches 1, a firm’s own price-elasticity, \(\epsilon_i(p_i, p_j)\), of demand becomes perfectly elastic. That is,

\[
\lim_{\beta \to 1} \epsilon_i(p_i, p_j) = \lim_{\beta \to 1} \frac{\frac{\partial q_i(p, p)}{\partial p_i} p}{q_i(p, p)} = \lim_{\beta \to 1} -\frac{p}{(1 - \beta)(\alpha - p)} = -\infty,
\]

implying a perfectly competitive outcome. Second, when prices are identical, the magnitude of a firm’s own price-elasticity of demand is monotonically increasing in \(\beta\). And third, when \(\beta\) approaches 0, a firm’s demand function becomes solely a function of its own price, implying that the firm behaves as a monopolist. These properties of \(\beta\) meet the traditional standards of “more competitive.”

Now moving to the firm’s profit maximization problem, denote \(\tilde{c}_i = c - x_i\) as firm \(i\)’s post-innovation marginal cost and let \(C(x_i) = \nu x_i^2 / 2\) be firm \(i\)’s cost for investing \(x_i\) units, where \(\nu > 0\) is an inverse measure of R&D efficiency. The net profit function for firm \(i\) is given by

\[
\Pi(p_i, p_j, x_i) = (p_i - \tilde{c}_i) \left( \frac{\alpha}{1 + \beta} - \frac{p_i - \beta p_j}{1 - \beta^2} \right) - \frac{\nu x_i^2}{2}, \quad i = 1, 2.
\]

Following Qiu (1997), the symmetric, sub-game perfect Nash equilibrium of this game
is

\[ x^* = \frac{2(\alpha - c)(2 - \beta^2)}{\nu(1 + \beta)(2 - \beta)(4 - \beta^2) - 2(2 - \beta^2)} \]
\[ p^* = \frac{\nu(\alpha - \alpha \beta + c)(1 + \beta)(4 - \beta^2) - 2\alpha(2 - \beta^2)}{\nu(1 + \beta)(2 - \beta)(4 - \beta^2) - 2(2 - \beta^2)} \]
\[ q^* = \frac{\nu(\alpha - c)(4 - \beta^2)}{\nu(1 + \beta)(2 - \beta)(4 - \beta^2) - 2(2 - \beta^2)} \]
\[ \Pi^* = 2(1 - \beta^2)q^* - \nu x^* \]
\[ z^* = \frac{\nu(x^*)^2/2}{p^*q^*} = \frac{2(\alpha - c)^2(2 - \beta^2)^2}{[\nu(\alpha - \alpha \beta + c)(1 + \beta)(4 - \beta^2) - 2\alpha(2 - \beta^2)]\nu(\alpha - c)(4 - \beta^2)} \]

where \( z^* \) is R&D intensity.

Qiu shows that a sufficient condition for the second-order and Routh-Hurwitz stability conditions to hold is \( \nu > \alpha/c \). Qiu also shows that \( \nu > \alpha/c \) is necessary and sufficient for post-innovation marginal costs to be positive. Imposing this assumption, I plot R&D intensity with respect to \( \beta \), holding \( \alpha \) and \( \nu \) fixed while varying \( c \). Fixing \( \alpha = 1 \) and \( \nu = 9/2 \), the figures corresponding to \( c = \{1/4, 1/3, 1/2\} \) (from left to right) are below:

The first graph is U-shaped with maximal R&D intensity occurring at maximal product substitutability. The second graph is also U-shaped, but maximal R&D intensity...
occurs with minimal product substitutability. And the third graph indicates that R&D intensity is monotonically decreasing in product substitutability. Thus, the relationship between R&D intensity and product substitutability in this model is not clear.

### 2.4 Measuring Innovation

Innovation is broadly defined as any change that adds value to an existing product, process or service or, as Peter F. Drucker defines it, [innovation] is the means by which the entrepreneur either creates new wealth-producing resources or endows existing resources with enhanced potential for creating wealth. It can range from a small change, such as a bump in computer processor speed, to a large one, such as a practical way to emit blue light through a light-emitting diode, or LED for short. Whatever its form, innovation is essential for economic progress. See the graphic below, for example; it plots (log) gross domestic product (GDP) per capita against the Global Innovation index by country.

---

14 Qiu does show that $x^*$ is monotonically decreasing in product substitutability for all other parameter values; but because greater product substitutability lowers gross profit in equilibrium and changes in investment are proportional to changes in gross profit, $x^*$ is not an informative measure of R&D effort.

15 See Drucker (1985).

16 Regarding blue LEDs, see Lincoln (2014). This innovation made it possible to produce white light from LEDs, the significance being that white LEDs “consume only about 5% of the power of an incandescent [bulb].”

17 Data are compiled using OECD statistics from stats.oecd.org and innovation metrics from the Global Innovation Index http://www.globalinnovationindex.org/content/page/data-analysis/.
From a stylized, theoretical modeling perspective, innovation is relatively straightforward to capture. We essentially need to account for the fact that innovation requires investment and that it generates value. To be more realistic, many models also account for the fact that the innovation process is inherently uncertain. I explore models that account for each of these aspects of innovation.

Choosing an empirical measure of innovation requires more care. In this case, the relationship between economic growth and innovation is key to determining an appropriate measure. However, as [Jorgenson (2011)](https://doi.org/10.1146/annurev-economics-020711-143751) argues, it is not enough:

Economic growth can take place without innovation through replication.

---

18 Innovation manifests, in part, as value-add. See [Economist (2015)](https://www.economist.com). The Economist discusses innovative activity in China and raises the issue of how to measure innovation. They echo the thoughts of the McKinsey Global Institute (MGI) regarding this matter: “MGI does not fall into the common trap of conflating innovation with invention: “The proof of successful innovation is the ability of companies to expand revenue and raise profits,” as opposed to filing lots of patents that never get used, or releasing a stream of novelty products that fail to generate a return.”
of established technologies. Investment increases the availability of these technologies, while the labor force expands as population grows. With only replication and without innovation, output will increase in proportion to capital and labor inputs, as suggested by Schultz (1956,1962). By contrast the successful introduction of new products and new or altered business models generates growth of output that exceeds the growth of capital and labor inputs.

Thus, an appropriate measure of innovation is one that, in addition to fulfilling the aspects of investment, value generation and uncertainty, corresponds to long-run economic growth. I follow Jorgenson and use total factor productivity as my measure of innovation. Other common measures of innovation include research and development expenditures and patents. However, both are accompanied with significant issues.

Patents are commonly used because of what they represent and how they are granted. More specifically, a patent should, in theory, only be granted if it is a sufficiently novel invention. However, a novel idea does not necessarily generate value. More importantly, patents reflect only a fraction of innovative output and may be the product of perverse economic incentives (Griliches (1990) and Hall and Harhoff (2012)).

Raw and citation-weighted patent counts are the two most common patent statistics used in the literature. Echoing what has already been said, raw patent counts are a poor proxy for innovation because: the incentives to patent go beyond the interest of protecting a novel and valuable idea (e.g., a company might apply for a patent to protect itself from potential litigation or to launch a lawsuit itself); the propensity to patent varies from entrepreneur to entrepreneur; and the ease with which a patent is granted depends on factors like the field of technology and what the patent examiner
considers to be novel.

Citation-weighted patent counts\(^{19}\) are better at capturing economic value, but they too suffer from significant bias. For example, patents with greater breadth (i.e., patents with more claims, broader language, etc.) are likely to receive a higher number of citations, all else equal. In addition, for any fixed window of time, some patents may receive more citations than others merely because they belong to a technology class characterized by high patenting activity. Recent research has actually shown evidence of an inverted-U relationship between citations and value (Abrams et al. (2013))—a clear violation of the assumption that more citations correspond to greater value. The explanation given for this relationship concerns the passiveness of the entrepreneur. In particular, for relatively low value patents, the entrepreneur is less concerned with protecting its position. Thus, the barrier to entry is relatively lax, giving way to more citations. On the other hand, for relatively high value patents, the entrepreneur is keen on protecting its competitive position, and thus works toward creating a patent that would make entry difficult.

Ultimately, patents represent a small portion of innovation, and evidence suggests that these innovations may be of relatively little value. For example, Cohen et al. (2000) survey the U.S. manufacturing sector and find that patents are the least effective at protecting intellectual assets. They find instead that the surveyed companies rely more heavily on trade secrecy and market lead time to protect their competitive position.

R&D intensity is another oft-used measure of innovation; it captures innovative effort. Being what it is, it cannot account for actual innovative output. This is a

\(^{19}\)Citation-weighted patent counts are constructed from forward citations, where a forward citation is a citation received from future patents.
clear drawback. Even as a measure of effort it can conceivably misidentify changes in effort. Consider the possibility of a firm switching from exploratory to exploitative R&D. Exploratory R&D is the type of R&D carried out to discover new ideas that may later be applied and brought to market. Exploitative R&D, in contrast, encompasses a firm developing already established research and taking it to market. Since exploratory R&D is more costly than exploitative R&D, we would observe an increase in expenditure if a firm switched from exploitative to exploratory R&D. At the same time, we would likely see an increase in sales, all else equal, if the ideas being developed and brought to market are successful. Altogether, a decrease in R&D intensity unrelated to a decrease in effort would follow.

R&D intensity can also decrease for budgetary and macroeconomic reasons, all else equal. For example, a firm might cut its R&D expenditures if it seeks to meet an internal budget or when the economy is in a downturn. This would reflect nothing more than cost-minimizing behavior.

I now turn attention to the development of a framework that allows us to analyze the competition-innovation relationship.

2.5 Toward a Robust Measure of Competition and its Relationship with Innovation

This section develops a game theoretic framework where the “contestability principle” of Shapiro (2012) is operative and firms optimally choose their innovation strategies given the threat their rival poses. Recall the definition of “more competitive”: an industry is “more competitive” if a firms stands to loser greater sales to its rivals

---

20 This example is borrowed from Mudambi et al. (2015).
should it offer inferior value to consumers. Two conditions are immediate from this definition. One, sales “switch hands” based on the value rivals offer to consumers. Most of the models I develop directly capture this by imposing a zero-sum property on competition. While this assumption may seem restrictive, it in fact allows us to separate the disciplinary effect of competition on innovation from other confounding factors. That is, if a firm innovates and subsequently sees an increase in its sales, the zero-sum property allows us to identify what part of the increase is due to a “business stealing” effect and what part of it is due to, perhaps, an increase in the size of the market. Two, a firm can only steal sales if it offers relatively greater value to consumers. This is explicit in each of my models. For example, in a game with stochastic innovation, I make the assumption that a firm can only steal profit from its rival if it is the sole innovator. The models now follow.

2.5.1 Model 1: Stochastic innovation with a linear penalty

This model examines the competition-innovation relationship when innovation is stochastic and when the gain to a sole innovator is linear in contestability. The model now follows.

Two firms compete with the expectation that, should only one successfully innovate, the unsuccessful firm will have a fraction of its profits stolen by its rival. If both firms succeed or both fail, each firm will maintain their competitive position. The expected change in profit for a firm is, therefore, an explicit function of relative success.

Let $P_i = p(x_i) \in [0, 1], i = 1, 2$, denote firm $i$’s probability of successful innovation

\[21\] Given the generality of the model, profits can be replaced with sales and the meaning would remain the same.
given its R&D effort, \( x_i \). Assume \( p' > 0 \) and \( p'' < 0 \). Let \( \phi \in (0,1) \) denote the fraction of profit stolen by a sole innovator. A larger \( \phi \) represents a more contestable market. Finally, assume that the cost of R&D effort is given by \( c_i = x_i \). The profit maximization problem facing firm \( i \), given its as well as its rival’s current level of profit, \( \pi_i \) and \( \pi_j \), is given by

\[
\mathbb{E}[\Pi_i|\pi_i,\pi_j] \equiv \max_{x_i \geq 0} \pi_i (1 - P_j)(\pi_i + \phi \pi_j) + (P_i P_j + (1 - P_i)(1 - P_j)) \pi_i \\
+ (1 - P_i)P_j (\pi_i - \phi \pi_i) - x_i \\
= \max_{x_i \geq 0} \pi_i + P_i (1 - P_j) \phi \pi_j - (1 - P_i)P_j \phi \pi_i - x_i,
\]

where \( \mathbb{E}[\cdot] \) is the expectation operator. The first term in (2.2) is firm \( i \)'s current profit; the second term is firm \( i \)'s expected gain in profit should it be the sole innovator; the third term is firm \( i \)'s expected loss in profit should its rival be the sole innovator; and the fourth term is the cost of firm \( i \)'s R&D effort. Note that the gain and loss functions for firm \( i \) are linear in the contestability parameter. That is, \( \phi \pi_j \) and \( \phi \pi_i \) are linear in \( \phi \). This will be relaxed in Model 2.

The first-order conditions of this game are given by

\[
P'_i = \frac{1}{\phi((1 - P_j)\pi_j + P_j \pi_i)}, \quad i = 1, 2, j \neq i,
\]

which implicitly define the best response functions, \( x_i = \omega_i(x_j, \phi) \), \( i = 1, 2 \), for each firm.

By the implicit function theorem, \( \omega_i \) will be a function of \( x_j \) and \( \phi \). In particular, firm \( i \)'s best response is given by \( x_i = \omega_i(x_j, \phi) \). It follows that firm \( i \)'s Nash equilibrium

---

22I analyze an example with a convex cost function. The numerical results are qualitatively the same.

23In this particular case, firm \( i \)'s, \( i = 1, 2 \), best response function will also be a function of \( \pi_i \) and \( \pi_j \). But for ease of notation, that is suppressed.
level of R&D effort, \( x_i^*(\phi) \), is implicitly given by
\[
x_i^*(\phi) \equiv \omega_i(x_i^*(\phi), \phi), \quad i = 1, 2, j \neq i. \tag{2.4}
\]

I assume that there exists an equilibrium strategy profile \((x_1^*, x_2^*)\) such that \( p(x_1^*) , p(x_2^*) \in [0,1] \) and \( \mathbb{E}[\Pi_i | x_i^*, x_j^*] \geq 0 \). Differentiating \(2.4\) with respect to \( \phi \) and solving for \( \frac{\partial x_i^*}{\partial \phi} \), we have
\[
\frac{\partial x_i^*}{\partial \phi} = \frac{\partial \omega_i(x_i^*, \phi)}{\partial x_j} \frac{\partial \omega_j(x_j^*, \phi)}{\partial \phi} - \frac{\partial \omega_i(x_i^*, \phi)}{\partial x_i} \frac{\partial \omega_j(x_j^*, \phi)}{\partial x_j} - \phi \frac{\partial \omega_i(x_i^*, \phi)}{\partial x_i} \tag{2.5}
\]

It will now be shown that aggregate R&D intensity, and therefore the aggregative probability of successful innovation, is strictly increasing with respect to contestability in equilibrium.

**Proposition 1** Given the assumptions \( p'(x) > 0 \) and \( p''(x) < 0 \) for all \( x > 0 \) and \( \phi \in (0,1) \), \( x^*(\phi) \equiv x_1^*(\phi) + x_2^*(\phi) \) is strictly increasing in \( \phi \). That is, aggregate R&D intensity is strictly increasing with respect to contestability.

**Proof 2.1** Implicit differentiation of \(2.3\) with respect to \( x_j \) and \( \phi \) gives
\[
\frac{\partial \omega_i}{\partial x_j} = -\frac{\phi P'_j (P'_i)^2 (\pi_i - \pi_j)}{P''_i}, \quad i = 1, 2, j \neq i
\]
\[
\frac{\partial \omega_i}{\partial \phi} = -\frac{P'_i}{\phi P''_i}, \quad i = 1, 2, j \neq i
\]

Evaluating these expressions at \( x_i = x_i^*, \ i = 1, 2, \) and plugging into \(2.5\) we have
\[
\frac{\partial x_i^*}{\partial \phi} = \frac{(P'_i P'_j)^2 (\pi_i - \pi_j)}{P''_i P''_j} - \frac{P'_i}{P''_i}, \quad i = 1, 2, j \neq i. \tag{2.6}
\]

Thus, the marginal change in aggregate R&D effort, \( x^* \equiv x_1^* + x_2^* \), is
\[
\frac{\partial x^*}{\partial \phi} = \frac{\partial x_i^*}{\partial \phi} + \frac{\partial x_j^*}{\partial \phi} = -\left( \frac{P'_i}{P''_i} + \frac{P'_j}{P''_j} \right) \frac{\phi}{\left[ 1 + \frac{\phi^2 (P'_i P'_j)^2 (\pi_i - \pi_j)^2}{P''_i P''_j} \right]}.
\]
Given the assumptions that $P_i' > 0$ and $P_i'' < 0$ for all $i$ and $\phi \in (0, 1)$, it is clear that both the numerator and denominator are positive. Thus, aggregate R&D effort is increasing in $\phi$. Q.E.D.

**Model 1: An example with a linear R&D cost function**

To capture the idea that the innovation is a “rare” event, assume the probability of successful innovation is given by $P_i = 1 - e^{-x_i}$, $i = 1, 2$; i.e., assume the outcome of innovation is Poisson distributed with an average rate of success $x_i$.

Following equation 2.3, the first-order conditions are given by

$$e^{-x_i} = \frac{1}{\phi(e^{-x_j}\pi_j + (1 - e^{-x_j})\pi_i)}, \quad i = 1, 2, j \neq i$$

or equivalently

$$x_i = \ln(\phi) + \ln \left( \pi_j - \pi_i + \pi_i e^{x_j} \right) - x_j, \quad i = 1, 2, j \neq i$$

Substituting firm $j$’s best response into firm $i$’s best response function and simplifying, we have

$$e^{2x^*_i} + \left( \frac{2(\pi_i - \pi_j) - \phi\pi_i\pi_j}{\pi_j} \right) e^{x^*_i} - \frac{\phi\pi_i(\pi_i - \pi_j)}{\pi_j} = 0,$$

where $x^*_i$ denotes the Nash equilibrium choice of $x_i$ for firm $i$. This is a quadratic equation in $e^{x_i}$. As such, we have two possible solutions:

$$x^*_i = \ln \left( \frac{z^i_2 - z^i_1 \pm \sqrt{(z^i_1)^2 + (z^i_2)^2}}{2} \right), \quad (2.7)$$

---

24A discrete random variable $Y$ is Poisson distributed if its density is

$$f(k; \lambda) = \Pr(Y = k) = \frac{\lambda^k e^{-\lambda}}{k!},$$

where $k$ is the number of successes during a unit interval of time, and $\lambda$ is the expected rate of success.

The probability of zero such events occurring, i.e., the probability of failure, is $\Pr(Y = 0) = e^{-\lambda}$. In turn, the probability of a success is $1 - e^{-\lambda}$. 
where \( z_1^i \equiv 2(\pi_i - \pi_j)/\pi_j \) and \( z_2^i \equiv \phi \pi_i \).

It will now be shown that each firm’s equilibrium R&D intensity is strictly increasing with respect to contestability.

**Lemma 2.1**

\[
x^*_i = \ln \left( \frac{z_2^i - z_1^i - \sqrt{(z_1^i)^2 + (z_2^i)^2}}{2} \right)
\]

is not a valid solution because it is never positive.

**Proof 2.2**

Consider

\[
x^*_i = \ln(g(z_1^i, z_2^i)),
\]

where

\[
g(z_1^i, z_2^i) \equiv \frac{z_2^i - z_1^i - \sqrt{(z_1^i)^2 + (z_2^i)^2}}{2}.
\]

The function \( g(\cdot) \) must be at least as large as 1 for \( x^*_i \) to be positive. Suppose this is the case. This holds if and only if

\[
\sqrt{(z_1^i)^2 + (z_2^i)^2} \leq -2 - (z_1^i - z_2^i).
\]

Since this inequality must hold by assumption and \( \sqrt{(z_1^i)^2 + (z_2^i)^2} \geq 0 \), the right-hand side must be non-negative, i.e., \( z_2^i \geq z_1^i + 2 \) must hold by necessity. Assume the latter holds, then. This means we can square both sides of the equation, implying that

\[
z_2^i(z_1^i + 2) \leq 2(z_1^i + 1)
\]

must hold. Since \( z_1^i \) can be negative or positive, \( z_1^i + 2 \leq 0 \). This implies two possible regions for \( z_2^i \):

\[
\left\{ z_1^i + 2 \leq z_2^i \leq \frac{2(z_1^i + 1)}{z_1^i + 2} : z_1^i \geq -2 \right\} \quad \text{or} \quad \left\{ z_2^i \geq \max \left\{ \frac{2(z_1^i + 1)}{z_1^i + 2}, z_1^i + 2 \right\} : z_1^i \leq -2 \right\}
\]

The first region is valid if and only if, for \( z_1^i \geq -2 \), \( 2(z_1^i + 1)/(z_1^i + 2) \geq z_1^i + 2 \) holds. But the latter holds if and only if \( h(z_1^i) \equiv (z_1^i)^2 + 2z_1^i + 2 \leq 0 \). Since \( h \) is quadratic in \( z_1^i \) and \( h'' > 0 \), the left-hand side of this inequality has a minimum, which occurs at
$z^i_1 = -1$. Evaluating at $z^i_1 = -1$, we arrive at $h(-1) = 1 > 0$. Thus, the first region is not valid.

This leaves us with the second region, which is non-empty. In fact, since $z^i_1 + 2 < 2(z^i_1 + 1)/(z^i_1 + 2)$ for all $z^i_1 \leq -2$, $z^i_2$ will lie above $2(z^i_1 + 1)/(z^i_1 + 2)$. The region where $z^i_2$ lies is consistent with the fact that $z^i_2 \geq 0$ must hold. However, $z^i_1 \leq -2$ implies

$$\frac{\pi_i}{\pi_j} \leq 0,$$

which violates the assumption that profits are strictly positive and bounded from above. We have reached a contradiction. Q.E.D.

**Lemma 2.2**

$x^*_i = \ln \left( \frac{z^i_2 - z^i_1 + \sqrt{(z^i_1)^2 + (z^i_2)^2}}{2} \right)$ is a valid solution.

**Proof 2.3** Using the definitions of $z^i_1$ and $z^i_2$ above, I first show that if firm $j$ loses a sufficiently large amount of profit to firm $i$ in the event that firm $i$ is the sole innovator, then $x^*_i \geq 0$ will hold trivially. However, if the amount of profit firm $i$ steals from firm $j$ is not too large, $x^*_i$ will be positive if the amount firm $i$ loses is sufficiently large.

Define

$$g(z^i_1, z^i_2) \equiv \frac{z^i_2 - z^i_1 + \sqrt{(z^i_1)^2 + (z^i_2)^2}}{2},$$

which must be at least as large as one for $x^*_i$ to be positive; i.e.,

$$\sqrt{(z^i_1)^2 + (z^i_2)^2} \geq 2 + z^i_1 - z^i_2 \quad (2.8)$$

must hold. Clearly, if the right-hand side is negative, this inequality will hold trivially. But the right-hand side is negative if and only if $\phi \pi_j > 2$. Thus, if firm $i$ steals a sufficiently large amount of profit from $j$, $x^*_i$ will be positive trivially.
Suppose, however, that the right-hand side is positive. Then (2.8) holds if and only if

\[(z_1^i)^2 + (z_2^i)^2 \geq 4 + 4(z_1^i - z_2^i) + (z_1^i)^2 + (z_2^i)^2 - 2z_1^iz_2^i \]

\[\iff z_2^i \geq \frac{2(z_1^i + 1)}{z_1^i + 2}, \text{ assuming } z_1^i + 2 > 0,\]

which implies that

\[\phi \pi_i \geq 2 - \frac{\pi_j}{\pi_i}\]

must hold. The right-hand side is obviously less than 2, so if \(\phi \pi_j < 2, \phi \pi_i > 2\) is a sufficient condition for a positive solution. Q.E.D.

**Proposition 2** \(x_i^* = \ln \left( \frac{z_2^i - z_1^i + \sqrt{(z_1^i)^2 + (z_2^i)^2}}{2} \right)\) is strictly increasing in \(\phi\). That is, regardless of firm status (leader vs. laggard), a firm’s equilibrium R&D intensity will increase in a more contestable market.

**Proof 2.4** The proof is trivial. Recall the definitions \(z_1^i = 2(\pi_i - \pi_j)/\pi_j\) and \(z_2^i = \phi \pi_i\) and that \(x_i^*\) may be rewritten as

\[x_i^* = \ln \left( \frac{z_2^i - z_1^i + \sqrt{(z_1^i)^2 + (z_2^i)^2}}{2} \right).\]

Clearly, \(z_1^i\) is independent of \(\phi\), so the effect of \(\phi\) on \(x_i^*\) happens only through \(z_2^i\). But it is obvious that \(x_i^*\) is strictly increasing in \(z_2^i\), and \(z_2^i\) is strictly increasing in \(\phi\). Thus, \(x_i^*\) must be strictly increasing in \(\phi\) by the chain-rule. Q.E.D.

The prediction that \(\partial x_i^*/\partial \phi > 0\) for \(i = 1, 2\) is a clearly a stronger result than aggregate R&D increasing with respect to \(\phi\). Intuitively, even if a firm has little to gain by innovating, the firm must also consider the risk of failing to innovate, which increases in a more contestable market. In turn, the firm will increase its R&D intensity in equilibrium when the market becomes more contestable.
Model 1: An example with a quadratic R&D cost function

So far only a linear R&D cost structure has been analyzed. In this section, I assume quadratic R&D investment costs. The example presented here does not yield an analytical solution. Consequently, I numerically solve the system of nonlinear reaction functions that are characterized by the first-order conditions.

The basic setup of the problem is the same as with the linear cost model, except now I assume \( c(x) = \nu x^2 / 2 \), where \( \nu > 0 \) represents R&D efficiency. The first-order conditions are given by

\[
e^{-x_i}e^{-x_j} \phi \pi_j + (1 - e^{-x_i})e^{-x_j} \phi \pi_i - \nu x_i = 0
\]
\[
e^{-x_j}e^{-x_i} \phi \pi_i + (1 - e^{-x_j})e^{-x_i} \phi \pi_j - \nu x_j = 0,
\]

which hold if and only if

\[
\nu x_i e^{x_i} = \phi \left( e^{-x_j} \pi_j + (1 - e^{-x_j}) \pi_i \right)
\]
\[
\nu x_j e^{x_j} = \phi \left( e^{-x_i} \pi_i + (1 - e^{-x_i}) \pi_j \right).
\]

Inverting each equation, we get the following reaction functions,

\[
x_i = W \left( \frac{\phi(e^{-x_j} \pi_j + (1 - e^{-x_j}) \pi_i)}{\nu} \right)
\]
\[
x_j = W \left( \frac{\phi(e^{-x_i} \pi_i + (1 - e^{-x_i}) \pi_j)}{\nu} \right),
\]

where \( W \) is the Lambert W function that defines the set of branches of the inverse relation \( f(x) = xe^x \). Letting \( s \equiv \pi_2 / (\pi_1 + \pi_2) \in [1/2, 1] \) and fixing \( \pi_2 \) to some constant \( \bar{\pi}_2 \), I numerically solve this system of nonlinear equations for all values of \( s \) and \( \phi \). Then, for every pair \((s, \phi) \in [1/2, 1] \times (0,1)\), I plot the average of equilibrium R&D intensities.\(^{25}\)

\(^{25}\)The search grids are \( s = [0.5, 0.55, 0.6, ..., 1] \) and \( \phi = [0, 0.01, 0.02, ..., 1] \). The fixed parameters are \( \pi_2 = \bar{\pi}_2 = 10000 \) and \( \nu = 30 \).
Inspection of the plot shows that average R&D intensity is increasing in $\phi$ for any fixed $s$. Likewise, for fixed $\phi$, average R&D intensity is decreasing in $s$. Thus, average R&D intensity reaches its maximum (minimum) when the industry is neck-and-neck (monopolized) and when competition is at its maximum (minimum).
2.5.2 Model 2: Stochastic innovation with a nonlinear penalty

In this next model, I examine the competition-innovation relationship under both deterministic and stochastic innovation. A more competitive market is, in principle, defined the same as in Model 1; however, now the gain and loss functions are nonlinear in $\phi$. The model now follows.

Two firms compete knowing that their profit is a direct function of relative value. Without loss of generality, assume firm 1 is the value leader and let $v \equiv v_1 - v_2 > 0$ denote the value “gap”, and $\Pi > 0$ total industry profit. The profit functions are given by:

$$
\pi(v; \phi) = \Pi f(\phi v) \\
\pi(-v; \phi) = \Pi f(-\phi v) = \Pi(1 - f(\phi v))
$$

where a larger $\phi > 0$ represents a more competitive market, and $f(\cdot)$ is a twice-continuously differentiable with the following properties:

I. $f(\phi v) \in [1/2, 1)$ and $f(-\phi v) \in (0, 1/2$

II. $f(\phi v) + f(-\phi v) = 1$

III. $f'(\phi v) > 0$ for all $v \geq 0$

IV. $f''(\phi v) < 0$ for all $v > 0$ and $f''(0) = 0$

V. $\lim_{v \to \infty} f(\phi v) = 1$ and $\lim_{v \to \infty} f(-\phi v) = 0$.

Thus, $f(\phi v)$ is an increasing, concave function in $v > 0$ with an upper limit of 1, whereas $f(-\phi v)$ is a decreasing, convex function in $v > 0$ with a lower limit of 0. The purpose of this type of function is to capture the “business stealing” effect of competition.
To see how a larger value of $\phi$ represents a more competitive market, first consider the case of a neck-and-neck industry, or $v = 0$. When the industry is neck-and-neck, the gain to a neck-and-neck firm from being a sole innovator is

$$g(0; \phi, \epsilon) \equiv \Pi(f(\phi(0 + \epsilon)) - f(0)) = \Pi(f(\phi \epsilon) - f(0)),$$

where $\epsilon > 0$ is the relative innovation size. Differentiating $g$ with respect to $\phi$, we have

$$\frac{\partial g}{\partial \phi} = \Pi \epsilon f'(\phi \epsilon) > 0.$$

Similarly, we have that the loss function, $l(0; \phi, \epsilon) \equiv \Pi(f(0) - f(-\phi \epsilon))$, increases with respect to $\phi$:

$$\frac{\partial l}{\partial \phi} = -\Pi \epsilon f'(-\phi \epsilon) > 0.$$

Thus, the gain and loss functions are increasing in $\phi$, the desired property of a more competitive market and the same as in Model 1. However, this property does not hold when the value gap is positive. Consider the gain to firm 1 should it be a sole innovator (the analysis is the same for the loss function). The gain to firm 1 is given by

$$g(v; \phi, \epsilon) = \Pi(f(\phi(v + \epsilon)) - f(\phi v)).$$

Differentiating with respect to $\phi$, we have

$$\frac{\partial g}{\partial \phi} = \Pi((v + \epsilon)f'(\phi(v + \epsilon)) - vf'(\phi v)).$$

Initially this will be positive, but for a sufficiently large $\phi$ it will be negative. Intuitively, $\phi$ accelerates profit growth, shifting firm 1’s profit share, $f(\phi v)$, outward. To get a clear sense of what this means, fix $v = v_0$ and suppose that $f(\phi v_0) = 60\%$. Increases in $\phi$ will push firm 1’s profit share closer and closer to $100\%$. Thus, at some point, we are effectively comparing firm 1’s gain at a relatively low profit share of
60% to its gain at a share close to 100%, which will be relatively small. This indicates that \( v \) is not the relevant state variable, profit share, \( s \), is. Or, to put it differently, \( v \) is the “nominal” gap and \( s \) is the “real” gap.

When we compare one competitive state to another via a change in \( \phi \), we need to hold \( s \) fixed, not \( v \). To this end, we need the value of \( v = \nu(s) \) such that \( s = f(\phi \nu(s)) \), or

\[
\nu(s) = \frac{f^{-1}(s)}{\phi}.
\]

Substituting \( v = \nu(s) \) into \( g(\cdot) \), we have

\[
g(\nu(s); \phi, \epsilon) = \Pi(f(\phi(\nu(s) + \epsilon)) - f(\phi \nu(s))) = \Pi(f(f^{-1}(s) + \phi \epsilon) - s).
\]

Notice now that the gain to firm 1 is increasing in \( \phi \):

\[
\frac{\partial g(\nu(s); \phi, \epsilon)}{\partial \phi} = \Pi \epsilon f'(f^{-1}(s) + \phi \epsilon) > 0.
\]

The parameter \( \phi \) is a means to an end; it is not “deep” in any sense. What ultimately matters for my analysis is that we have parameter that, when larger, mirrors a more contestable market. And this will be case for any parameter that induces an increase in the gain and loss functions regardless of profit or market share.

**Deterministic Innovation**

Now that we have established that larger \( \phi \) corresponds to a more contestable market, consider the following problem where both firms choose their levels of value to maximize their net profit. Let \( v_i \) denote firm \( i \)'s value, and suppose the cost of innovation is given by \( c(v_i) = 1/2v_i^2 \). Firm 1 and 2 solve the following:

\[
\max_{v_1} \Pi f(\phi(\nu(s) + v_1 - v_2)) - \frac{1}{2} v_1^2 \\
\max_{v_2} \Pi f(-\phi(\nu(s) + v_1 - v_2)) - \frac{1}{2} v_2^2
\]
Notice that the profit share of firm $i$ can only go up if it generates more value than its rival; i.e., $v_i > v_j$ must attain. This reflects one of the key aspects of the definition of “more competitive,” namely that a firm can only capture profitable sales from its rival if it offers *superior* value. The first-order conditions are:

$$
\Pi \phi f' (\phi (\nu (s) + v_1 - v_2)) - v_1 = 0
$$

$$
\Pi \phi f' (-\phi (\nu (s) + v_1 - v_2)) - v_2 = 0
$$

Since $f' (\phi (\nu (s) + v_1 - v_2)) = f' (-\phi (\nu (s) + v_1 - v_2))$ by assumption (II), the equilibrium choices will be identical, implying an optimal value $v^*_1 = v^*_2 = v^*$

$$
v^*(\Pi, \phi, s) = \Pi \phi f(f^{-1}(s)).
$$

Thus, both firms will increase their value in response to an increase in $\phi$. In addition to this result, we find that a neck-and-neck industry, i.e., $s = 1/2$, maximizes innovation:

$$
\frac{\partial v^*}{\partial s} = \Pi \phi f''(f^{-1})'
$$

$$
= \frac{\Pi \phi f''}{f'} < 0, \quad \text{by assumption (III) and (IV) and the Inverse Function Theorem}
$$

Because $s \in [1/2, 1)$, the last line implies that innovation is maximized when firms are neck-and-neck. Caution should be taken with this result since the shape of $f(\cdot)$ matters. For example, if the leader’s profit growth is sigmoidal, then innovation will be maximized at the point of inflection. Indeed, one will find that innovation initially increases with profit share, but then declines.

---

26Note that the second-order conditions for a maximum are satisfied trivially since $f''(\cdot) < 0$
Stochastic Innovation

This next model assumes innovation is stochastic. Let \( P_i \in [0, 1] \), \( i = 1, 2 \), denote firm \( i \)'s probability of successful innovation, and let \( \eta > 0 \) denote R&D efficiency. Firm \( i \) solves the following:

\[
\begin{align*}
\max_{P_i \in [0,1]} & \quad P_i (1 - P_j)(\pi_i + g_i) + (P_i P_j + (1 - P_i)(1 - P_j))\pi_i \\
& + (1 - P_i)P_j (\pi_i - l_i) - \eta P_i^2 / 2 \\
= & \max_{P_i \in [0,1]} \pi_i + P_i (1 - P_j)g_i - (1 - P_i)P_j l_i - \eta P_i^2 / 2,
\end{align*}
\]

where

\[
\begin{align*}
\pi_i & \equiv \pi(\nu(s_i); \phi) = \Pi s_i \\
g_i & \equiv g(\nu(s_i); \phi, \epsilon) = \Pi (f(f^{-1}(s_i) + \phi \epsilon) - s_i) \\
l_i & \equiv l(\nu(s_i); \phi, \epsilon) = \Pi (s_i - f(f^{-1}(s_i) - \phi \epsilon))
\end{align*}
\]

The first-order condition of this problem is

\[
(1 - P_j)g_i + P_j l_i - \eta P_i = 0,
\]

implying that

\[
P_i = \frac{1}{\eta}((1 - P_j)g_i + P_j l_i)
\]

must hold in equilibrium. It is instructive to briefly discuss the interpretation of this first-order condition. It says that the equilibrium probability firm \( i \) chooses will be proportional to its incentive to innovate. To see this, note that the incentive to innovate is defined as the difference between expected profit conditional on successfully innovating and expected profit conditional on failing to innovate. Let \( \iota_i \) be an indicator variable that equals 1 if firm \( i \) successfully innovates and 0 otherwise. Then, firm
i’s incentive to innovate is given by

\[
E[\tilde{\pi}_i | \xi = 1] - E[\tilde{\pi}_i | \xi = 0] = P_j\pi_i + (1 - P_j)(\pi_i + g_i)
\]

\[
- (P_j(\pi_i - l_i) + (1 - P_j)\pi_i)
\]

\[
= (1 - P_j)g_i + P_jl_i
\]

where \(\tilde{\pi}_i\) is a firm i’s random profit. Thus, in the absence of strategic effects, firm i will unambiguously choose a higher probability of success in a more contestable market.

For ease of notation, let \(g\) and \(l\) define firm 1’s gain and loss, respectively, and recall that firm 1 is the leader. The equilibrium probabilities are then given by:

\[
P_1^* = \frac{\eta g + l(l - g)}{\eta^2 + (l - g)^2}
\]

\[
P_2^* = \frac{\eta l - g(l - g)}{\eta^2 + (l - g)^2}
\]

Observe that \(l > g\) since \(g\) is concave in \(s\). As a result, the equilibrium probability of success for a leader is trivially positive, while for a laggard it is not. To ensure that firm 2’s probability is positive, we require that

\[
\eta > \frac{g}{l}(l - g)
\]

hold.

Observe now that the aggregate probability of success is given by

\[
P^* \equiv P_1^* + P_2^* = \frac{\nu(l + g) + (l - g)^2}{\nu^2 + (l - g)^2}
\]

It will be shown that the aggregate probability of success is strictly increasing with respect to a more contestable market.

**Lemma 2.3**

\[
g_\phi \equiv \partial g / \partial \phi < l_\phi \equiv \partial l / \partial \phi
\]
Proof 2.5 Observe that
\[ l_\phi - g_\phi = \Pi e [f'(f^{-1}(s) - \phi e) - f'(f^{-1}(s) + \phi e)] > 0, \text{ by concavity of } f \text{ for } s > 1/2. \]
Q.E.D.

Next observe that the partial effect of \( \phi \) on \( P^* \) is given by
\[
\frac{\partial P^*}{\partial \phi} = \frac{\frac{\partial P^*_1}{\partial \phi} + \frac{\partial P^*_1}{\partial \phi} \frac{\partial P^*_2}{\partial \phi} + \frac{\partial P^*_2}{\partial \phi} \frac{\partial P^*_1}{\partial \phi}}{1 - \frac{\partial P^*_1}{\partial \phi} \frac{\partial P^*_2}{\partial \phi}} 
\bigg[ - \frac{(1 - P^*_2)g_\phi + P^*_2 l_\phi}{\eta} + \frac{(l - g)((1 - P^*_1)l_\phi + P^*_1 g_\phi)}{\eta^2} + \frac{(1 - P^*_1)l_\phi + P^*_1 g_\phi}{\eta} + \frac{(g - l)((1 - P^*_2)g_\phi + P^*_2 l_\phi)}{\eta^2} \bigg] 
\bigg[ 1 + \frac{(l - g)^2}{\eta^2} \bigg] 
\bigg[ \frac{\eta((1 - P^*_2)g_\phi + P^*_2 l_\phi) + (l - g)((1 - P^*_1)l_\phi + P^*_1 g_\phi)}{\eta^2 + (l - g)^2} 
+ \frac{\eta((1 - P^*_1)l_\phi + P^*_1 g_\phi) + (g - l)((1 - P^*_2)g_\phi + P^*_2 l_\phi)}{\eta^2 + (l - g)^2} \bigg],
\]
where I use the definition of a Nash equilibrium for the first line. Let \( x \equiv (1 - P^*_1)l_\phi + P^*_1 g_\phi > 0 \) and \( y \equiv (1 - P^*_2)g_\phi + P^*_2 l_\phi > 0 \), so that \( \partial P^*/\partial \phi \) may be rewritten as
\[
\frac{\partial P^*}{\partial \phi} = \frac{(\eta + l - g)x + (\eta + g - l)y}{\eta^2 + (l - g)^2}.
\]
Thus,
\[
\frac{\partial P^*}{\partial \phi} < 0 \iff (\eta + l - g)x < (l - g - \eta)y.
\]
If \( \eta > l - g, \partial P^*/\partial \phi < 0 \) will not hold.

Proposition 3 For \( \eta \) sufficiently large such that \( P^*_1 < 1 \) for all \( (\phi, s) \in (0, \infty) \times [1/2, 1) \), it must be the case that \( \partial P^*/\partial \phi > 0 \) holds.

Suppose that \( \frac{\eta}{\eta^2 + (l - g)^2} < \eta < 1 - g \), a necessary but not sufficient condition for \( \partial P^*/\partial \phi < 0 \) to hold and which implies \( P^*_1 > P^*_2 > 0 \). Because \( P^*_1 > P^*_2 \), ensuring \( P^*_1 < 1 \) will likewise ensure that \( P^*_2 < 1 \). Recall that
\[
P^*_1 = \frac{\eta g + l(l - g)}{\eta^2 + (l - g)^2}.
\]
This is less than 1 if and only if

\[ F_1(\eta) \equiv \eta^2 - \eta g - g(l - g) > 0. \]

Note that \( F_1 \) is strictly increasing in \( \eta \) for \( \eta > g/2 \). Setting \( F_1 = 0 \) and solving for \( \eta \), we arrive at

\[ \hat{\eta} = \frac{1}{2} \left( g + \sqrt{g^2 + 4g(l - g)} \right). \]

It is trivial to show that \( \hat{\eta} > g/2 \); thus, for \( F_1 > 0 \) to hold, we require that \( \eta > \hat{\eta} \).

Observe that \( \hat{\eta} \) is a strictly increasing function in \( \phi \) (via the fact that \( g \) and \( l \) are strictly increasing in \( \phi \) and \( l_\phi - g_\phi > 0 \)). Thus, for \( P_1^* \) to hold for all \( \phi \), take the limit of \( \hat{\eta} \) as \( \phi \to \infty \) and note that

\[ \lim_{\phi \to \infty} g = \Pi(1 - s) \]
\[ \lim_{\phi \to \infty} l = \Pi s \]

Thus,

\[ \lim_{\phi \to \infty} \hat{\eta} = \frac{1}{2} \left( \Pi(1 - s) + \sqrt{\Pi^2(1 - s)^2 + 4\Pi^2(1 - s)(2s - 1)} \right) \]

Differentiating with respect to \( s \) and setting equal to zero, we find that the value of \( s \) which maximizes \( \lim_{\phi \to \infty} \hat{\eta} \) is

\[ \hat{s} = \frac{1}{7} \left( 5 - \frac{\sqrt{2}}{2} \right) \approx 0.6133 \]

Plugging \( s = \hat{s} \) into \( \lim_{\phi \to \infty} \hat{\eta} \), we have

\[ \eta > \frac{\Pi}{7} \left( 1 + 2\sqrt{2} \right) \quad (2.9) \]

This is a necessary and sufficient condition for \( P_1^* < 1 \) to hold for all \( (\phi, s) \in (0, \infty) \times (1/2, 1] \).
The next step is to find the condition that guarantees $\partial P^*/\partial \phi < 0$. This will attain whenever
\[
\frac{2(l - g)(l + g - \eta)}{\eta^2 + (l - g)^2} > \frac{l_\phi + g_\phi}{l_\phi - g_\phi} > 1.
\]
Thus, at the very least
\[
F_2(\eta) \equiv \eta^2 + 2(l - g)\eta + (l - g)^2 - 2(l - g)(l + g) < 0
\]
must be satisfied for $\partial P^*/\partial \phi < 0$ to attain.\footnote{This does not guarantee that $\partial P^*/\partial \phi < 0$. However, by examining this weaker condition, we reach a stronger result. To see this, let $z \equiv (l_\phi + g_\phi)/(l_\phi - g_\phi) > 1$ and define
\[
F_2^*(\eta) \equiv \eta^2 + \frac{2(l - g)}{z} \eta + (l - g)^2 - \frac{2(l - g)(l + g)}{z},
\]
which must be negative for $\partial P^*/\partial \phi < 0$ to hold. Observe that
\[
F_2^*(\eta) - F_2(\eta) > 0 \iff \eta < l + g.
\]
By assumption, $\eta < l - g$, so $F_2^*$ lies above $F_2$. And since both $F_2$ and $F_2^*$ are increasing in $\eta$, it must be the case that the value of $\eta$ which satisfies $F_2 = 0$ will be greater than the value of $\eta$ which satisfies $F_2^* = 0$.}
the limit of $\tilde{\eta}$ as $\phi \to \infty$, we have

$$
\lim_{\phi \to \infty} \tilde{\eta} = -\Pi(2s - 1) + \Pi \sqrt{2(2s - 1)}
$$

Differentiating with respect to $s$ and setting equal to zero, we find that the value of $s$ which maximizes $\lim_{\phi \to \infty} \tilde{\eta}$ is

$$
\tilde{s} = \frac{3}{4}
$$

Plugging $s = \tilde{s}$ into $\lim_{\phi \to \infty} \tilde{\eta}$, we have

$$
\eta < \frac{\Pi}{2}
$$

(2.10)

This is a necessary but not sufficient condition for $\partial P^*/\partial \phi < 0$ to attain. However, the upper-bound on $\eta$ given by 2.10 is less than the lower-bound on $\eta$ given by 2.9 since $1 + 2\sqrt{2} > 7/2$. Consequently, $\partial P^*/\partial \phi < 0$ can only attain when $P_1^* > 1$ and $\eta < l - g$. Q.E.D.

A sufficient condition for $\partial P^*/\partial \phi > 0$ is $\eta > l - g$. Thus, by choosing $\eta$ large enough, there exists valid equilibria $(P_1^*, P_2^*) \in (0, 1) \times (0, 1)$ such that $\partial P^*/\partial \phi > 0$.

The model outlined has just shown that, regardless of strategic effects, the total probability of successful innovation increases in response to a more competitive market.

2.5.3 Model 3: Stochastic process innovation in a two-stage game of strategic R&D and price competition with cost-asymmetric firms

An industry is composed of two cost-asymmetric firms that price compete via a differentiated good. The product space is defined over the unit interval $[0, 1]$. A unit mass of consumers with uniformly distributed preferences lie along this space, each with unit demand. The location of a firm is taken as exogenous, with firm 1 located at 0 and firm 2 located at 1.
The indirect utility consumer $c$ receives by purchasing a good from firm $i$, $i = 1, 2$, is:

$$U_c = r - p_i - \frac{1}{\phi}D(L_c, L_i),$$

where $r$ is the maximum willingness to pay for the good, $p_i$ is the price offered by firm $i$, $1/\phi$ is the unit cost a consumer incurs for choosing a good outside its preference and $D(L_c, L_i) = (L_c - L_i)^2$ is the distance between a consumer located at $L_c$ and a firm located at $L_i \in \{0, 1\}$. The marginal consumer, i.e., the one indifferent between the offerings of firm 1 and 2, is found by equating

$$r - p_1 - \frac{1}{\phi}\hat{L}_c^2 = r - p_2 - \frac{1}{\phi}(\hat{L}_c - 1)^2$$

and solving for $\hat{L}_c$. Doing so, we find

$$\hat{L}_c = (1 - \phi(p_1 - p_2))/2,$$

implying that the demand curves facing firm 1 and 2 are:

$$d_1(p_1, p_2) = \int_0^{\hat{L}_c} 1 \cdot dc = \hat{L}_c = (1 - \phi(p_1 - p_2))/2$$

$$d_2(p_1, p_2) = 1 - \int_0^{\hat{L}_c} 1 \cdot dc = 1 - \hat{L}_c = (1 - \phi(p_2 - p_1))/2.$$

In this context of price competition, the parameter $\phi$ is consistent with “more competitive;” the extent to which a firm can capture sales from its rival by offering a relatively lower price, and therefore relatively greater value, is governed by $\phi$. Intuitively, a larger $\phi$ corresponds to a lower switching, or transportation, cost, and this will magnify the loss from offering a relatively higher price. Note that $\phi$ is sometimes referred to as a product substitutability parameter. This is not technically correct.
since zero product substitutability implies monopoly—an implication that does not arise here, as both firms split the market when $\phi$ equals zero.

Moving to the firm’s profit maximization problem, assume both firms have a constant marginal cost of production, $c_i$. Then firm $i$ solves:

$$\pi_i(p_i, p_j) = \max_{p_i} (p_i - c_i)(1 + \phi(p_j - p_i))/2 \quad i = 1, 2.$$ 

The first-order conditions are given by

$$1 + \phi(p_j - p_i) - \phi(p_i - c_i) = 0 \quad i = 1, 2.$$

Solving for the Nash equilibrium of this game, we get

$$p_i^* = \phi^{-1} + \frac{2c_i + c_j}{3} \quad i = 1, 2.$$

Substituting $p_i^*, i = 1, 2,$ into $\pi_i(p_i, p_j)$ gives

$$\pi_i(\Delta_i) = \frac{1}{2} \left( \frac{1}{\phi} - \frac{\Delta_i}{3} \right) \left( 1 - \frac{\phi\Delta_i}{3} \right) = \frac{2}{\phi}(d_i^*)^2, \quad i = 1, 2,$$

where $\Delta_i \equiv c_i - c_j$.

The objective is to now investigate a stochastic game of process innovation similar to Model 1 and Model 2. To this end, I impose zero-sum competition by assuming firms vie for profit share. Assume without loss of generality that firm 1 is the cost leader, i.e., $\Delta \equiv c_2 - c_1 > 0$. Firm $i$’s profit share is given by

$$S(\phi\Delta_i) \equiv \frac{\pi_i(\Delta_i)}{\pi_i(\Delta_i) + \pi_j(-\Delta_i)} = \frac{d_i^2}{d_i^2 + d_j^2} = \frac{d_i^2}{d_i^2 + (1 - d_i)^2} = \frac{1}{2} \frac{(3 - \phi\Delta_i)^2}{9 + \phi^2\Delta_i^2} \quad i = 1, 2, j \neq i.$$

The relationship between profit share and the efficiency gap is as expected: we find that a larger $\phi$ generates a larger loss for a neck-and-neck firm should it offer inferior value, i.e., if $\Delta > 0$ (see the graph below):
This graph plots the relatively inefficient firm’s profit share against the efficiency gap, \(\Delta > 0\). When \(\phi\) is larger, the relatively inefficient firm’s profit share reaches zero more rapidly. This is precisely how \(\phi\) affected profit share in Model 2. Consequently, the “nominal” state is \(\Delta\) and the “real” state is profit share, \(s\).

The value of \(\Delta > 0\), \(\delta(s)\), such that \(S(\delta(s)) = s\) is

\[
\delta(s) = \frac{3}{\phi} \frac{2\sqrt{s(1-s)} - 1}{2s - 1} = \frac{z(s)}{\phi}, \quad i = 1, 2,
\]

where

\[
z(s) \equiv \frac{3(2\sqrt{s(1-s)} - 1)}{2s - 1}.
\]

Now define the gain, \(g\), and loss, \(l\), functions that arise from innovation:

\[
g(\delta(s); \phi, \epsilon) \equiv S(z(s) - \phi \epsilon) - s
\]

\[
l(\delta(s); \phi, \epsilon) \equiv s - S(z(s) + \phi \epsilon)
\]

These functions are increasing in \(\phi\), making \(\phi\) an appropriate measure of competition. First consider the case of a neck-and-neck industry, \(i.e., s = 1/2\). The gain and loss functions are given by:

\[
g(0; \phi, \epsilon) = \frac{1}{2} \frac{(3 + \phi \epsilon)^2}{9 + \phi^2 \epsilon^2} - \frac{1}{2}
\]

\[
l(0; \phi, \epsilon) = \frac{1}{2} - \frac{1}{2} \frac{(3 - \phi \epsilon)^2}{9 + \phi \epsilon^2}
\]
Differentiating both the gain and loss with respect to $\phi$, we have

$$\frac{\partial g(0; \phi, \epsilon)}{\partial \phi} = \frac{\partial l(0; \phi, \epsilon)}{\partial \phi} = \frac{3\epsilon(9 - \phi^2 \epsilon^2)}{(9 + \phi^2 \epsilon^2)^2},$$

which is non-negative for all $\phi \in [0, 3/\epsilon]$. The upper bound on $\phi$ is not restrictive because it corresponds to the state where the market is captured entirely by one firm. More specifically, the profit share of the post-innovation leader cannot exceed 1, or

$$\frac{1}{2} \frac{(3 + \phi \epsilon)^2}{9 + \phi^2 \epsilon^2} \leq 1 \iff \phi \leq \frac{3}{\epsilon}.$$

As a result, both the prospective gain and loss for a neck-and-neck firm is strictly increasing over the relevant range of $\phi$.

Consider now the general case for any $s$, and recall that

$$S(\phi \Delta_i) = \frac{d_i^2}{d_i^2 + (1 - d_i)^2}.$$

In general,

$$\frac{\partial S(\phi \Delta_i)}{\partial \phi} = \frac{2 \frac{\partial d_i}{\partial \phi} d_i (1 - d_i)}{(d_i^2 + (1 - d_i)^2)^2} = \frac{\partial d_i}{\partial \phi} \frac{2d_i (1 - d_i)}{(1 - 2d_i (1 - d_i))^2}.$$

The sign of $\partial S(\phi \Delta_i)/\partial \phi$ is therefore the same as $\partial d_i/\partial \phi$. That is,

$$\text{sgn} \left( \frac{\partial g(\delta(s); \phi, \epsilon)}{\partial \phi} \right) = \text{sgn} \left( \frac{\partial d_i(\delta(s) - \epsilon)}{\partial \phi} \right),$$

$$\text{sgn} \left( \frac{\partial l(\delta(s); \phi, \epsilon)}{\partial \phi} \right) = -\text{sgn} \left( \frac{\partial d_i(\delta(s) + \epsilon)}{\partial \phi} \right).$$

If a firm $i$ is a sole innovator, then demand for firm $i$'s product will become

$$d_i(\delta(s) - \epsilon) = \frac{1}{2} \left( 1 - \frac{\phi(\delta(s) - \epsilon)}{3} \right)$$

$$= \frac{1}{2} \left( 1 - \frac{\phi \left( \frac{z(s)}{\phi} - \epsilon \right)}{3} \right)$$

$$= \frac{1}{2} \left( 1 - \frac{z(s) - \phi \epsilon}{3} \right),$$
Thus, we have $\frac{\partial d_i(\delta(s) - \epsilon)}{\partial \phi} = \epsilon / 6 > 0$. Applying the same analysis, it is straightforward to show that if firm $i$’s rival is a sole innovator, an increase in $\phi$ will amplify its loss by $\epsilon / 6$. The gain and loss functions are consequently increasing in $\phi$ regardless of firm status.

The model, which is fundamentally the same as Model 2, now follows. Let $P_i \in [0, 1], i = 1, 2$, denote firm $i$’s probability of successful innovation, and let $\eta > 0$ denote R&D efficiency. Firm $i$ solves the following:

$$\max_{P_i \in [0, 1]} P_i (1 - P_j) (S_i + g_i) + (P_i P_j + (1 - P_i) (1 - P_j)) S_i$$

$$+ (1 - P_i) P_j (S_i - l_i) - \eta P_i^2 / 2$$

$$= \max_{P_i \in [0, 1]} S_i + P_i (1 - P_j) g_i - (1 - P_i) P_j l_i - \eta P_i^2 / 2,$$

where

$$S_i \equiv S(\delta(s_i); \phi) = \frac{d_i^2}{d_i^2 + (1 - d_i)^2}$$

$$g_i \equiv g(\delta(s_i); \phi, \epsilon) = S(z(s_i) - \phi \epsilon) - s_i$$

$$l_i \equiv l(\delta(s_i); \phi, \epsilon) = s_i - S(z(s_i) + \phi \epsilon)$$

The first-order condition of this problem is

$$(1 - P_j) g_i + P_j l_i - \eta P_i = 0,$$

implying that

$$P_i = \frac{1}{\eta} ((1 - P_j) g_i + P_j l_i)$$

must hold in equilibrium.

The Nash equilibrium of the game is characterized by the first-order conditions. For ease of notation, let $g$ and $l$ define firm 1’s gain and loss, respectively, and recall
that firm 1 is the leader. The equilibrium probabilities are then given by:

\[ P_1^* = \frac{\eta g + l(l - g)}{\eta^2 + (l - g)^2} \]
\[ P_2^* = \frac{\eta l - g(l - g)}{\eta^2 + (l - g)^2} \]

Observe that \( l > g \) since \( g \) is concave in \( s \). As a result, the equilibrium probability of success for a leader is trivially positive, while for a laggard it is not. To ensure that firm 2’s probability is positive, we require that

\[ \eta > \frac{g}{l}(l - g) \]

hold. The total probability of success is

\[ P^* \equiv P_1^* + P_2^* = \frac{\nu(l + g) + (l - g)^2}{\nu^2 + (l - g)^2} \]

Like in Model 2, it can be shown that for sufficiently large \( \eta \) such that \( P_1^* < 1 \) for all \((\phi, s)\), \( \partial P^*/\partial \phi > 0 \) will hold for all \((\phi, s)\). The proof is omitted.

2.5.4 Model 4a: Deterministic process innovation in a two-stage game of strategic R&D and price competition with cost-symmetric firms

An industry is composed of two symmetric firms, indexed by \( i = 1, 2 \), where innovation and price decisions are made in a two-stage game. This game considers process innovation, where firms invest to lower their marginal costs of production.

The preference-space, defined over the unit interval \([0, 1]\), is characterized by a unit mass of consumers with uniformly distributed preferences. Each consumer has unit demand. The location of a firm is taken as exogenous, with one located at 0 (firm \( i \)) and the other at 1 (firm \( j \)).

The indirect utility a consumer receives from firm \( i \) is given by:

\[ U = r - p_i - \frac{1}{\phi} d(L_c, 0)^2, \]
where \( p_i \) is the price paid, \( d(L_c,0)^2 = (L_c - L_i)^2 \) is the distance between a consumer located at \( L_c \) and firm \( i \) located at 0, and \( \phi \) is a parameter that reduces consumer transportation costs. The marginal consumer, \( i.e., \) the one indifferent between the offerings of firm \( i \) and \( j \), is found by equating

\[
s - p_i - \frac{1}{\phi} \hat{L}_c^2 = s - p_j - \frac{1}{\phi} (1 - \hat{L}_c)^2.
\]

Solving, we get

\[
\hat{L}_c = \frac{1}{2} - \frac{\phi}{2} (p_i - p_j),
\]

implying that the demand curves facing firm \( i \) and \( j \) are:

\[
D_i(p) = \hat{L}_c - 0 = \frac{1}{2} - \frac{\phi}{2} (p_i - p_j)
\]

\[
D_j(p) = 1 - \hat{L}_c = \frac{1}{2} - \frac{\phi}{2} (p_j - p_i),
\]

where \( p = (p_1, p_2) \). Both firms have the same constant marginal cost of production, \( c \). The cost of investing \( x \) units to lower marginal cost to \( \tilde{c} = c - x \) is \((1/2)x^2\).

The profit maximization problem firm \( i \) faces is:

\[
V_i(p) = \max_{p_i, x_i} (p_i - \tilde{c})D_i(p), \tag{2.11}
\]

It will now be shown that in a sub-game perfect Nash equilibrium, R&D is strictly increasing with respect to contestability.

We first begin with the second stage, where firm \( i, i = 1, 2 \), chooses \( p_i \) given \( \{p_j, x_i\} \) to maximize its gross profit, \( i.e., \) it solves

\[
\max_{p_i} (p_i - \tilde{c})D_i(p) - \frac{1}{2} x_i^2
\]

The corresponding first-order conditions are

\[
\frac{1}{2} (1 - \phi (p_i - p_j)) - \frac{\phi}{2} (p_i - \tilde{c}) = 0, \quad i = 1, 2
\]
This system of equations implies the following best-response price functions in terms of \( x = (x_1, x_2) \),

\[
p_i(x) = \phi^{-1} + c - \frac{2x_i + x_j}{3}
\]

Substituting these into 2.11 firm \( i \) then solves

\[
\max_{x_i} (p_i(x) - \tilde{c})D_i(p_i(x), p_j(x)) - \frac{1}{2}x_i^2
\]

to maximize its net profit. The first-order conditions are

\[
\frac{1}{3} (1 - \phi(x_j - x_i) - x_i) = 0, \ i = 1, 2,
\]

implying the equilibrium levels of R&D intensity, defined as the ratio of R&D expenditure to gross profit, are

\[
r_i^* = r_j^* = r^* = \frac{\phi}{9}.
\]

Thus, in a symmetric, subgame perfect Nash equilibrium, R&D intensity increases the more contestable the market becomes.
2.5.5 Model 4b: Deterministic process innovation in a two-stage game of strategic R&D and price competition with cost-asymmetric firms

An industry is composed of two firms where process innovation and price decisions are made in a two-stage game. Process innovation is done in the first stage to lower marginal costs, and then prices are chosen in the second stage to maximize gross profit. How effective a firm is at creating or maintaining demand for its product will depend on its relative price offering and the ease with which consumers can switch between products.

The product space is defined over the unit interval \([0, 1]\). A unit mass of consumers with uniformly distributed preferences lie along this space. Each consumer has unit demand. The location of a firm is taken as exogenous, with firm 1 located at 0 and firm 2 located at 1.

The indirect utility a consumer receives by purchasing a good from firm \(i\):

\[
U_c = r - p_i - \frac{1}{\phi} D(L_c, L_i),
\]

where \(r\) is the maximum willingness to pay for the good, \(p_i\) is the price offered by firm \(i, i = 1, 2\), \(1/\phi\) is the unit cost a consumer incurs for choosing a good outside its preference and \(D(L_c, L_i) = (L_c - L_i)^2\) is the distance between a consumer located at \(L_c\) and a firm located at \(L_i\). The marginal consumer, i.e., the one indifferent between the offerings of firm 1 and 2, is found by equating

\[
r - p_1 - \frac{1}{\phi} \hat{L}_c^2 = r - p_2 - \frac{1}{\phi} (\hat{L}_c - 1)^2
\]

and solving for \(\hat{L}_c\). Doing so, we find

\[
\hat{L}_c = \frac{(1 - \phi(p_1 - p_2))}{2},
\]
implying that the demand curves facing firm 1 and 2 are:

\[ d_1(p_1, p_2) = \int_0^{i_c} 1 \cdot dc = \hat{L}_c = (1 - \phi(p_1 - p_2))/2 \]

\[ d_2(p_1, p_2) = 1 - \int_0^{i_c} 1 \cdot dc = 1 - \hat{L}_c = (1 - \phi(p_2 - p_1))/2. \]  \(2.12\)

The firms are asymmetric in terms of their constant marginal costs of production, \(c_i, i = 1, 2\). Without loss of generality, assume \(\Delta_c \equiv c_2 - c_1 > 0\). To offer a lower price while still remaining profitable, each firm makes an investment, \(x_i, i = 1, 2\), toward lowering its marginal cost, where the cost of investment is given by \(c(x) = x^2/2\). Thus, firm \(i\)'s post-innovation marginal cost is \(\tilde{c}_i = c_i - x_i\).

In this context, the parameter \(\phi\) measures contestability because a larger value of \(\phi\) results in a larger zero-sum loss in demand should a firm offer relatively inferior value, \(i.e., a relatively higher price. It is standard, however, to view \(\phi\) as a consumer transportation cost in spatial competition models, \(i.e., as capturing how costly it is for a consumer to switch from one firm’s good to the other. In this respect, larger \(\phi\) lowers the cost of switching, which may ultimately may be interpreted as making the market more contestable.

Firm \(i\)'s gross and net profit are

\[ \pi_i(p_i, p_j, x_i) = (p_i - \tilde{c}_i)(1 + \phi(p_j - p_i))/2 \]  \(2.13\)

\[ \Pi_i(p_i, p_j, x_i) = (p_i - \tilde{c}_i)(1 + \phi(p_j - p_i))/2 - x_i^2/2, \]  \(2.14\)

The solution we seek is a set of subgame perfect Nash equilibrium investment and price strategies. Thus, the game is solved via backward induction. Starting with the second stage, firms choose their price strategies simultaneously. The first-order
The first-order conditions corresponding to the first stage are

\[
\frac{\partial \Pi_1}{\partial x_1} = \frac{(1 + \phi(\Delta_c + x_1 - x_2)/3)}{3} - x_1 = 0 \tag{2.19}
\]

\[
\frac{\partial \Pi_2}{\partial x_2} = \frac{(1 - \phi(\Delta_c + x_1 - x_2)/3)}{3} - x_2 = 0 \tag{2.20}
\]

implying the reaction functions

\[
x_1(x_2) = \frac{3 + \phi(\Delta_c - x_2)}{9 - \phi}
\]

\[
x_2(x_1) = \frac{3 - \phi(\Delta_c + x_1)}{9 - \phi} \tag{2.21}
\]
Solving (2.21) we find that the optimal investment levels are

\[ x_1^* = \frac{1}{3} \left( 1 + \frac{3\phi \Delta_c}{B} \right) \]
\[ x_2^* = \frac{1}{3} \left( 1 - \frac{3\phi \Delta_c}{B} \right) \]

where \( B \equiv 9 - 2\phi \). The equilibrium prices, quantities, net profits and gross profits now follow. Substituting (2.22) into (2.16) we get equilibrium prices

\[ p_1^* = \phi^{-1} + \frac{1}{3} \left( 2c_1 + c_2 - 1 - \frac{\phi \Delta_c}{B} \right) \]
\[ p_2^* = \phi^{-1} + \frac{1}{3} \left( 2c_2 + c_1 - 1 + \frac{\phi \Delta_c}{B} \right) \] (2.23)

To get equilibrium quantities, substitute (2.23) and (2.22) into (2.12):

\[ d_1^* = \frac{1}{2} \left( 1 + \frac{3\phi \Delta_c}{B} \right) = \frac{3x_1^*}{2} \]
\[ d_2^* = \frac{1}{2} \left( 1 - \frac{3\phi \Delta_c}{B} \right) = \frac{3x_2^*}{2} \] (2.24)

Finally, substituting (2.22) into (2.17) and (2.18) gives us equilibrium gross and net profits, respectively:

\[ \pi_1^* = 2\phi^{-1}(d_1^*)^2 \] (2.25)
\[ \Pi_1^* = \frac{2(B + \phi)}{9\phi} (d_1^*)^2 \] (2.26)
\[ \pi_2^* = 2\phi^{-1}(d_2^*)^2 \] (2.27)
\[ \Pi_2^* = \frac{2(B + \phi)}{9\phi} (d_2^*)^2 \] (2.28)

The existence of a feasible equilibrium will now be established, as will the monotonically positive relationship between R&D intensity and contestability.

Lemma 2.4

Satisfaction of the Routh-Hurwitz stability condition implies satisfaction of the second-order condition
Proof 2.6

\[
\frac{\partial^2 \Pi_i}{\partial x_i^2} \frac{\partial^2 \Pi_j}{\partial x_j^2} - \frac{\partial^2 \Pi_i}{\partial x_j \partial x_i} \frac{\partial^2 \Pi_i}{\partial x_i \partial x_j} = \frac{9 - 2\phi}{9} = \frac{B}{9} > 0, \quad i = 1, 2, j \neq i
\]

\[
\frac{\partial^2 \Pi_i}{\partial x_i^2} = \frac{\phi - 9}{9} = -\frac{B + \phi}{9} < 0, \quad i = 1, 2.
\]

It is clear that the Routh-Hurwitz stability condition holds if and only if \( B > 0 \). And since \( \phi > 0 \) by assumption, the second-order condition for a maximum also holds. Q.E.D.

Lemma 2.5

A sufficient condition for positive post-innovation marginal costs, \( c_i - x_i^* \), \( i = 1, 2 \), is that the leader’s post-innovation marginal cost, \( c_1 - x_1^* \), be positive.

Proof 2.7 The post-innovation marginal costs of the leader and laggard are

\[
c_1 - x_1^* = \frac{9c_1 - \phi(c_1 + c_2)}{B} - \frac{1}{3}
\]

\[
c_2 - x_2^* = \frac{9c_2 - \phi(c_1 + c_2)}{B} - \frac{1}{3}
\]

Since \( c_2 - x_2^* - (c_1 - x_1^*) = 9\Delta_c/B > 0 \), we need only have the leader’s post-innovation marginal cost be positive to ensure that both post-innovation costs are positive. Q.E.D.

Lemma 2.6

If \( c_1 + c_2 > 2/3 \), then the leader’s post-innovation marginal cost, \( c_1 - x_1^* \), is positive if and only if \( \phi < 9(3c_1 - 1)/(3(c_1 + c_2) - 2) \). Otherwise, it is positive if and only if \( \phi > 9(3c_1 - 1)/(3(c_1 + c_2) - 2) \).

Proof 2.8 The leader’s post-innovation marginal cost is positive if and only if

\[
\phi(3(c_1 + c_2) - 2) < 9(3c_1 - 1).
\]
This implies two candidate regions for $\phi$:

Region 1: $\phi < 9(3c_1 - 1)/(3(c_1 + c_2) - 2)$, if $3(c_1 + c_2) - 2 > 0$

Region 2: $\phi > 9(3c_1 - 1)/(3(c_1 + c_2) - 2)$, if $3(c_1 + c_2) - 2 < 0$

Q.E.D.

Lemma 2.7

Given the restriction that $\phi > 0$, Region 1 from Lemma 2 is valid only if $c_1 > 1/3$. If we further restrict $c_1 < 2/3$, then the laggard’s innovation level, $x_2^*$, will be positive under Region 1. On the other hand, Region 2 from Lemma 2 is valid for all $c_1 > 0$; however, if $c_1 < 1/3$, Region 2 implies that the laggard’s equilibrium innovation level, $x_2^*$, is negative.

Proof 2.9 It is clear that if $c_1 > 1/3$, Region 1 is a valid candidate region for $\phi$. Now suppose $c_1 \in (1/3, 2/3)$. This implies positive innovation from the laggard in equilibrium. To see this, observe that the laggard’s equilibrium innovation level is positive if and only $\phi < 9/(3(c_2 - c_1) + 2)$. This upper bound is greater than the one under Region 1 if and only if

$$\frac{9}{3(c_2 - c_1) + 2} > \frac{9(3c_1 - 1)}{3(c_1 + c_2) - 2},$$

which holds if and only if $c_1 < 2/3$. Thus, if $c_1 \in (1/3, 2/3)$, Region 1 is sufficient to ensure that the laggard’s innovation level is positive in equilibrium.$^{28}$

Similarly, it is clear that Region 2 is a valid candidate for all $c_1 > 0$. Region 2, however, is uninformative if $c_1 > 1/3$, since this just implies that $\phi$ must be greater than a negative number. Suppose then that $c_1 < 1/3$. This implies that the laggard’s

$^{28}$Note that the leader’s innovation level will be positive for all $\phi, \Delta c, B > 0$, so no additional analysis is needed for the leader’s innovation level.
equilibrium innovation level is negative for all $\phi$ in this region. To see this, suppose
the upper bound of this region is greater than the lower bound implied by Region 2
when $c_1 < 1/3$; i.e., suppose
\[
\frac{9}{3(c_2 - c_1) + 2} > \frac{9(3c_1 - 1)}{3(c_1 + c_2) - 2}.
\]
Simple algebra shows that this holds if and only if $c_1 > 2/3$; but $c_1 < 1/3$ and
$c_1 + c_2 < 2/3$ by assumption. Thus, the laggard’s equilibrium innovation level will be
negative in Region 2. Q.E.D.

Lemma 2.8

Given that $c_1 \in (1/3, 2/3)$, a sufficient condition for positive equilibrium innovation
levels, post-innovation marginals costs, prices, quantities, gross profits and net profits
is that $\phi$ be bounded from above according to Region 1 of Lemma 2.

Proof 2.10 We have already proved that Region 1 is sufficient for positive post-
innovation marginal costs and innovation levels. Now observe that each firm’s equi-
librium demand is proportional to its own innovation level, and that gross and net
profits are quadratic in demand. Thus, equilibrium quantities and gross and net
profits must also be positive under Region 1 when $c_1 \in (1/3, 2/3)$.

We now need to show that prices will be positive under Region 1. Note that
$p_1^* - p_2^* = -\Delta_c(1 + 2\phi/B) < 0$ for all values of the parameters; we therefore only
need to show that $p_1^*$ is positive under Region 1. Now observe that $p_1^*$ is strictly
decreasing in $\phi$, since
\[
\frac{\partial p_1^*}{\partial \phi} = -\phi^{-2} - \frac{9\Delta_c}{3B^2} < 0.
\]
Thus, if by evaluating $p_1^*$ at the upper bound of Region 1 we find that $p_1^*$ is still
positive, then both the leader’s and laggard’s prices will be positive in equilibrium.
To this end, we find that

\[
p^*_1 \left( \phi = \frac{9(3c_1 - 1)}{3(c_1 + c_2) - 2} \right) = \frac{1}{3} \frac{c_1(3(c_1 + c_2) - 2)}{3c_1 - 1},
\]

(2.29)

which is positive since \( c_1 > 1/3 \) and \( c_1 + c_2 > 2/3 \) by assumption. Q.E.D.

Henceforth, attention will be restricted to Region 1 for values of \( c_1 \in (1/3, 2/3) \).

Proposition 4 Average prices and the Herfindahl-Hirschman Index are strictly decreasing and increasing, respectively, with respect to \( \phi \).

Proof 2.11 Average prices and the Herfindahl-Hirschman Index are given by

\[
P \equiv \frac{(p^*_1 + p^*_2)}{2} = \phi^{-1} + \frac{3(c_1 + c_2) - 2}{6}
\]

\[
H \equiv (d^*_1)^2 + (d^*_2)^2 = \frac{1}{2} \left( 1 + \left( \frac{3\phi \Delta c}{B} \right)^2 \right).
\]

Inspection alone shows that the average price and the Herfindahl-Hirschman Index are decreasing and increasing, respectively. Q.E.D.

Proposition 5 R&D intensity in terms of gross and net profit is strictly increasing in \( \phi \). Thus, average R&D intensity in terms of either gross or net profit is strictly increasing in \( \phi \).

Proof 2.12 For both the leader and laggard, R&D intensity in terms of gross and net profit is given by

\[
r^g_i \equiv \frac{(x_i)^2/2}{\pi^*_i} = \frac{(x^*_i)^2/2}{2\phi^{-1} \cdot 9(x^*_i)^2/4} = \frac{\phi}{9}, \quad i = 1, 2
\]

\[
r^n_i \equiv \frac{c(x_i)}{\Pi^*_i} = \frac{(x^*_i)^2/2}{2(9 - \phi)/9\phi^{-1} \cdot 9(x^*_i)^2/4} = \frac{\phi}{9 - \phi}, \quad i = 1, 2
\]

Inspection shows that both R&D intensity in terms of gross profit \( (r^g_i) \) and net profit \( (r^n_i) \) is strictly increasing in \( \phi \). Q.E.D.

\[\text{Note that the average price is unambiguously positive since } 3(c_1 + c_2) - 2 > 0.\]
2.5.6 Model 5: Deterministic product innovation in a two-stage game of strategic R&D and price competition with cost-asymmetric firms

The previous model explored the effect of contestability on process innovation when the only strategic variable was price. This next model focuses on price and quality competition, where investment is made to increase quality, not reduce costs.

An industry is composed of two firms where product innovation and price decisions are made in a two-stage game. Product innovation is done in the first stage to shift demand up, and then prices are chosen in the second stage to maximize gross profit. How effective a firm is at creating or maintaining demand for its product will depend on its relative price and quality offerings and the ease with which consumers can switch between products.

The product space is defined over the unit interval $[0, 1]$. A unit mass of consumers with uniformly distributed preferences lie along this space. Each consumer has unit demand. The location of a firm is taken as exogenous, with firm 1 located at 0 and firm 2 located at 1.

The indirect utility a consumer receives by purchasing a good from firm $i$:

$$U_c = r + (x_i - p_i) - \frac{1}{\phi}D(L_c, L_i),$$

where $r$ is the maximum willingness to pay for the good, $x_i$ and $p_i$ are the respective quality and price offered by firm $i$, $i = 1, 2$, $1/\phi$ is the unit cost a consumer incurs for choosing a good outside its preference and $D(L_c, L_i) = (L_c - L_i)^2$ is the distance between a consumer located at $L_c$ and a firm located at $L_i$. The marginal consumer, i.e., the one indifferent between the offerings of firm 1 and 2, is found by equating

$$r + (x_1 - p_1) - \frac{1}{\phi}\hat{L}_c^2 = r + (x_2 - p_2) - \frac{1}{\phi}(\hat{L}_c - 1)^2$$
and solving for $\hat{L}_c$. Doing so, we find

$$\hat{L}_c = (1 - \phi((x_2 - p_2) - (x_1 - p_1))) / 2,$$

implying that the demand curves facing firm 1 and 2 are:

$$d_1(p_1, p_2, x_1, x_2) = \int_0^{L_c} 1 \cdot dc = \hat{L}_c = (1 - \phi((x_2 - p_2) - (x_1 - p_1))) / 2$$

$$d_2(p_1, p_2, x_1, x_2) = 1 - \int_0^{L_c} 1 \cdot dc = 1 - \hat{L}_c = (1 - \phi((x_1 - p_1) - (x_2 - p_2))) / 2.$$

(2.30)

I define the difference between quality and price, $x_i - p_i$, as value. Hence, demand is a direct function of relative value.

The firms are asymmetric in terms of their constant marginal costs of production, $c_i$, $i = 1, 2$. Without loss of generality, assume $\Delta_c \equiv c_2 - c_1 > 0$. To shift demand up, each firm makes an investment, $x_i$, $i = 1, 2$, towards increasing its quality, where the cost of investment is given by $c(x) = x^2 / 2$.

In this context, the parameter $\phi$ measures contestability because a larger value of $\phi$ results in a larger zero-sum loss in demand should a firm offer relatively inferior value, i.e., a relatively smaller quality-price difference $x - p$.

Firm $i$’s gross and net profit are

$$\pi_i(p_i, p_j, x_i, x_j) = (p_i - c_i)(1 - \phi((x_j - p_j) - (x_i - p_i))) / 2$$

$$\Pi_i(p_i, p_j, x_i, x_j) = (p_i - c_i)(1 - \phi((x_i - p_i) - (x_j - p_j))) / 2 - x_i^2 / 2,$$

(2.31) (2.32)

The solution we seek is a set of subgame perfect Nash equilibrium investment and price strategies. Thus, the game is solved via backward induction. Starting with the second stage, firms choose their price strategies simultaneously. The first-order
conditions are

\[
\frac{\partial \pi_1}{\partial p_1} = \frac{1 - \phi(x_2 - p_2 - (x_1 - p_1) + (p_1 - c_1))}{2} = 0
\]

\[
\frac{\partial \pi_2}{\partial p_2} = \frac{1 - \phi(x_1 - p_1 - (x_2 - p_2) + (p_2 - c_2))}{2} = 0
\]

implying the reaction functions

\[
p_1(p_2, x_1, x_2) = \frac{(p_2 + c_1 + x_1 - x_2 + \phi^{-1})}{2}
\]

\[
p_2(p_1, x_1, x_2) = \frac{(p_1 + c_2 + x_2 - x_1 + \phi^{-1})}{2}
\] (2.33)

Solving the system of equations 2.33, we get

\[
p_1(x_1, x_2) = \phi^{-1} + \frac{(2c_1 + (2c_2 + c_1 + x_2 - x_1))}{3}
\]

\[
p_2(x_1, x_2) = \phi^{-1} + \frac{(2c_2 + c_1 + x_2 - x_1)}{3}
\] (2.34)

We now substitute 2.34 into 2.31 and 2.32 to give us firm i’s, i = 1, 2, gross and net profit, respectively, as a function only of innovation strategies:

\[
\pi_i(x_i, x_j) = \frac{(\phi^{-1} - (c_i - c_j + x_j - x_i)/3)(1 - \phi(c_i - c_j + x_j - x_i))}{2}
\] (2.35)

\[
\Pi_i(x_i, x_j) = \frac{(\phi^{-1} - (c_i - c_j + x_j - x_i)/3)(1 - \phi(c_i - c_j + x_j - x_i))}{2} - x_i^2/2
\] (2.36)

The first-order conditions corresponding to the first stage are

\[
\frac{\partial \Pi_1}{\partial x_1} = \frac{(1 - \phi(x_2 - x_1 - \Delta_c)/3)}{3} - x_1 = 0
\] (2.37)

\[
\frac{\partial \Pi_2}{\partial x_2} = \frac{(1 - \phi(\Delta_c + x_1 - x_2)/3)}{3} - x_2 = 0
\] (2.38)

implying the reaction functions

\[
x_1(x_2) = \frac{3 + \phi(\Delta_c - x_2)}{9 - \phi}
\]

\[
x_2(x_1) = \frac{3 - \phi(\Delta_c + x_1)}{9 - \phi}
\] (2.39)
Solving (2.39) we find that the optimal investment levels are

\[
x_1^* = \frac{1}{3} \left( 1 + \frac{3\phi \Delta_c}{B} \right),
\]
\[
x_2^* = \frac{1}{3} \left( 1 - \frac{3\phi \Delta_c}{B} \right),
\]

where \( B \equiv 9 - 2\phi \). The equilibrium prices, quantities, net profits and gross profits now follow. Substituting (2.40) into (2.34) we get equilibrium prices

\[
p_1^* = \phi^{-1} + \frac{1}{3} \left( 2c_1 + c_2 + \frac{2\phi \Delta_c}{B} \right)
\]
\[
p_2^* = \phi^{-1} + \frac{1}{3} \left( 2c_2 + c_1 - \frac{2\phi \Delta_c}{B} \right)
\]

To get equilibrium quantities, substitute (2.41) and (2.40) into (2.30)

\[
d_1^* = \frac{1}{2} \left( 1 + \frac{3\phi \Delta_c}{B} \right) = \frac{3x_1^*}{2}
\]
\[
d_2^* = \frac{1}{2} \left( 1 - \frac{3\phi \Delta_c}{B} \right) = \frac{3x_2^*}{2}
\]

Finally, substituting (2.40) into (2.35) and (2.36) gives us equilibrium gross and net profits, respectively:

\[
\pi_1^* = 2\phi^{-1}(d_1^*)^2
\]
\[
\Pi_1^* = \frac{2(B + \phi)}{9\phi} (d_1^*)^2
\]
\[
\pi_2^* = 2\phi^{-1}(d_2^*)^2
\]
\[
\Pi_2^* = \frac{2(B + \phi)}{9\phi} (d_2^*)^2
\]

The existence of a feasible equilibrium will now be established, as will the monotonically positive relationship between R&D intensity and contestability.

**Lemma 2.9**

Satisfaction of the Routh-Hurwitz stability condition implies satisfaction of the second-order condition.
Proof 2.13
\[ \frac{\partial^2 \Pi_i}{\partial x_i^2} \frac{\partial^2 \Pi_j}{\partial x_j^2} = \frac{\partial^2 \Pi_i}{\partial x_j \partial x_i \partial x_i \partial x_j} = \frac{9 - 2\phi}{9} = \frac{B}{9} > 0, \quad i = 1, 2, j \neq i \]
\[ \frac{\partial^2 \Pi_i}{\partial x_i^2} = \frac{\phi - 9}{9} = \frac{-B + \phi}{9} < 0, \quad i = 1, 2. \]

It is clear that the Routh-Hurwitz stability condition holds if and only if \( B > 0 \). And since \( \phi > 0 \) by assumption, the second-order condition for a maximum also holds. Q.E.D.

Lemma 2.10
Assume \( \phi < \min \left\{ \frac{1}{2} \left( -\frac{1}{\Delta_c} + \sqrt{\frac{1}{\Delta_c^2} + \frac{18}{\Delta_c}} \right), \frac{9}{3\Delta_c+2} \right\} \). This is a sufficient condition to ensure that quantities, prices, investment levels, gross profits, and net profits are positive in equilibrium. Moreover, it is a necessary condition to ensure that even if the laggard were to hypothetically choose the same price as the leader in the second stage, it would still have positive market share.

Proof 2.14 It is clear that the leader’s innovation level is always positive, but the laggard’s can be negative. Thus, we restrict attention to the laggard’s innovation level, which is positive if and only if \( \phi < 9/(3\Delta_c+2) \). Likewise, because the laggard’s innovation is always less than the leader’s, it is hypothetically possible that the laggard’s market share could be negative were it to choose a price at least as large as the leader in the second stage (in other words, the laggard’s demand intercept could be negative). That is, \( 1/2(1 - \phi(x_1 - x_2)) \) could be negative. To ensure that this is positive, we must have that \( \phi < \frac{1}{2} \left( -\frac{1}{\Delta_c} + \sqrt{\frac{1}{\Delta_c^2} + \frac{18}{\Delta_c}} \right) \). It is straightforward to show that the bound on \( \phi \) corresponding to positive market share is initially less than the bound on \( \phi \) corresponding to a positive innovation level, but eventually becomes greater. Hence, to guarantee positive innovation and market share, we require that \( \phi < \min \left\{ \frac{1}{2} \left( -\frac{1}{\Delta_c} + \sqrt{\frac{1}{\Delta_c^2} + \frac{18}{\Delta_c}} \right), \frac{9}{3\Delta_c+2} \right\} \). Q.E.D.
Lemma 2.11

Prices, quantities and profits are positive in equilibrium under Lemma 1.10.

Proof 2.15 Observe that demand is proportional to innovation levels, and profits are quadratic in demand. Thus, if innovation is positive—which the bound on $\phi$ guarantees—then demand and profits will be positive.

Notice next that the price of the leader is positive for all parameter values and the price of the laggard is strictly decreasing in $\phi$. Therefore, we need only inspect the laggard’s price to prove that the our condition ensures positive prices. Because the laggard’s price is strictly decreasing in $\phi$, and $\phi < \min \left\{ \frac{1}{2} \left( -\frac{1}{\Delta c} + \sqrt{\frac{1}{\Delta c^2} + \frac{18}{\Delta}} \right); \frac{9}{3\Delta c + 2} \right\}$, we need only show that $p_2^*$ evaluated at one of the bounds on $\phi$ is positive. And evaluating $p_2^*$ at the bound of $\phi$ corresponding to positive innovation, we find that $p_2^* = c_2$. Q.E.D.

Proposition 6 The average price and the Herfindahl-Hirschman Index are strictly decreasing and increasing, respectively, with respect to $\phi$.

Proof 2.16 The average price and the Herfindahl-Hirschman Index are given by

\[ \bar{P} \equiv (p_1^* + p_2^*)/2 = \phi^{-1} + \frac{c_1 + c_2}{2}, \]

\[ H \equiv (d_1^*)^2 + (d_2^*)^2 = \frac{1}{2} \left( 1 + \left( \frac{3\phi \Delta c}{B} \right)^2 \right). \]

Inspection alone shows that the average price and the Herfindahl-Hirschman Index are decreasing and increasing, respectively. Q.E.D.

Proposition 7 R&D intensity in terms of gross and net profit is strictly increasing in $\phi$. Thus, average R&D intensity in terms of either gross or net profit is strictly increasing in $\phi$. 
Proof 2.17 For both the leader and laggard, R&D intensity in terms of gross and net profit is given by

\[ r^g_i \equiv \frac{c(x_i)}{\pi_i^*} = \frac{(x_i^*)^2/2}{2\phi^{-1} \cdot 9(x_i^*)^2/4} = \frac{\phi}{9}, \ i = 1, 2 \]

\[ r^n_i \equiv \frac{c(x_i)}{\Pi_i^*} = \frac{(x_i^*)^2/2}{2(9 - \phi)/9\phi^{-1} \cdot 9(x_i^*)^2/4} = \frac{\phi}{9 - \phi}, \ i = 1, 2 \]

Inspection shows that both R&D intensity in terms of gross profit \( r^g_i \) and net profit \( r^n_i \) is strictly increasing in \( \phi \). Q.E.D.

2.6 Applying the model to an empirical framework

Each of the models I have explored predicts a positive relationship between contestability and innovation. To test this prediction empirically, I follow Model 1, where innovation is stochastic and firms steal profitable sales from one another if there is a sole innovator. This model offers the purest, most simple representation of the contestability principle.

Let \( s \) denote sales and \( v \) consumer value. Then the representation of \( \phi \) in Model 1 may, in principle, may be approximated via the following equation:

\[ s = \phi v + \epsilon, \quad (2.47) \]

where \( \epsilon \) is an \( iid \), zero-mean error term. To see this, suppose firm \( i \) offers greater value than firm \( j, j \neq i \), i.e., \( v_i > v_j \). Then

\[ \mathbb{E}[s_i - s_j|v] = \phi(v_i - v_j). \quad (2.48) \]

Thus, the amount of sales firm \( j \) could have had in expectation, \( \mathbb{E}[s_i - s_j|v] \), but did not because it offered less value than firm \( i \), is magnified by the parameter \( \phi \). Unlike the theoretical model, however, where rivals actually steal sales from one another, equation \( 2.48 \) operates at a hypothetical level.
Estimation of 2.47 requires data on consumer value, which is unobservable. There is no way around this. However, assume the following holds:

\[ m = \beta s + \eta, \tag{2.49} \]

where \( m \) is market value/capitalization and \( \eta \) is an iid, zero-mean error term uncorrelated with \( \epsilon \). Then 2.47 and 2.49 imply

\[ m = \beta \phi v + (\beta \epsilon + \eta). \]

Thus, the correlation between market value and sales gives us an implicit way to estimate contestability.

Moving to the competition-innovation equation, let \( y \equiv f(\beta) \) denote innovation, where \( f(\cdot) \) is some smooth, twice-continuously differentiable function and \( \beta \) is the extent of contestability captured in equation 2.49. Taking a second-order Taylor series expansion around \( \beta = \beta_0 \), we have

\[
y = f(\beta_0) + f'(\beta_0)(\beta - \beta_0) + \frac{f''(\beta_0)}{2}(\beta - \beta_0)^2 + \nu
\]

\[
= \left[ f(\beta_0) - f'(\beta_0)\beta_0 + \frac{f''(\beta_0)}{2}\beta_0^2 \right] + (f'(\beta_0) - f''(\beta_0))\beta + \frac{f''(\beta_0)}{2}\beta^2 + \nu
\]

\[
= \alpha_0 + \alpha_1 \beta + \alpha_2 \beta^2 + \nu, \tag{2.50}
\]

where \( \alpha_0 \equiv f(\beta_0) - f'(\beta_0)\beta_0 + \frac{f''(\beta_0)}{2}\beta_0^2 \), \( \alpha_1 \equiv f'(\beta_0) - f''(\beta_0) \), \( \alpha_2 \equiv \frac{f''(\beta_0)}{2} \) and \( \nu \) is the error term due to approximation. Abstracting away from other factors that predict innovation, equations 2.49 and 2.50 are the basis of my empirical strategy.

We now turn to Chapter 3 which develops the empirical model.
Chapter 3

Competition and Innovation: An Empirical Test

3.1 Introduction

The question of whether or not market power spurs innovation has been a source of debate among economists ever since Schumpeter (1942) dismissed perfect competition as the ideal model for economic growth. On the one hand, Schumpeter rejects perfect competition as an ideal worth striving for because it fails to account for the dynamic nature of capitalism; it does not consider the introduction of new goods, services, and methods of production and business. On the other hand, he argues that perfect competition is incompatible with innovation because entrants would promptly erode the economic profit of would-be innovators. The latter is the primary source of contention in the literature on market power and innovation. Another source of contention is his position on “big business.” Schumpeter points out that large firms may be better at facilitating innovation via their access to capital and their ability to absorb the risk inherent in the innovation process.

The total of Schumpeter’s argument has lead to a substantial empirical literature. Beginning in the 1960s and ending in the 1980s, the literature focused on firm size and market concentration and their effect on research and development (R&D) activity. Scherer (1967), for example, investigates the relationship between the fraction of R&D personnel and market concentration in the cross section. He finds that the
fraction of R&D personnel initially increases with market concentration, but then declines. Overall, however, this literature presents weak evidence of an economically and statistically significant relationship (Gilbert (2006)).

A new wave of literature emerged in the 1990s, focusing instead on the relationship between competition and innovation. Nickell (1996) was the first to investigate the competition-innovation relationship. Under a linear specification, he finds that competition, measured principally in terms of average rents normalized on value-added, is positively correlated with productivity growth. Blundell et al. (1999) also estimate a linear model. Measuring innovation in terms of innovation counts and competition in terms of market share, concentration and important penetration, they likewise find a positive correlation. Following these studies, a trend toward to non-linear models developed. This trend continues today.

Non-linear models came into favor following the work of Aghion et al. (2005). Based on a panel of publicly traded manufacturing firms in the United Kingdom during the period 1973-1994, they estimate a Poisson maximum likelihood model and find empirical evidence of an inverted-U relationship at the industry level with citation-weighted patent counts as their measure of innovation and operating margins as their measure of competition. Their work has generated substantial interest in the literature. Notable examples include Correa (2012), Correa and Ornaghi (2014) and Hashmi (2013), all of which are derivative works of Aghion et al. (2005).

Correa (2012) uses the same data and baseline model as Aghion et al. (2005) and finds that the inverted-U breaks down when allowing for structural breaks. The possibility of a structural break is motivated by his hypothesis that the establishment of the United States Court of Appeals for the Federal Circuit in 1982 had a significant impact on patenting behavior. His parametric and non-parametric tests for a struc-
tural break confirm his hypothesis. The parametric test of a structural break in 1982 cannot be rejected at the 5% significance level, and the non-parametric test points to a statistically significant break in 1981. Taking these structural breaks into account, he finds that the competition-innovation relationship is positive between 1973-1982, but not statistically significant during the period 1983-1994.

Correa and Ornaghi (2014) revisit the model of Aghion et al. (2005) with a different sample of data, namely a panel of publicly traded manufacturing firms in the United States over a similar time horizon, and two independent modifications. One modification is with respect to the distributional assumption on citation-weighted patent counts. Instead of assuming a Poisson distribution, they assume a negative binomial. The assumption of a Poisson distribution is relaxed because it requires equivalence between the mean and variance of patent counts. This assumption is not met in their data, meaning Poisson maximum likelihood estimation would result in incorrect standard errors. They consequently perform a negative binomial regression—which leaves the mean function unchanged—to avoid the problem of “overdispersion.” The other modification consists of using productivity as a measure of innovation. In both cases, they find evidence of monotonic and positive relationship.

Hashmi (2013) also revisits Aghion et al. (2005). Like Correa and Ornaghi (2014), he performs a negative binomial regression of citation-weighted patent counts on operating margins with a panel of publicly traded manufacturing firms in the United States. The time horizon of his data also closely matches Correa and Ornaghi (2014). However, unlike Correa and Ornaghi (2014), he finds a “mildly” negative relationship. He reconciles his result with that of Aghion et al. (2005) via a theoretical model. Specifically, his model predicts an interaction between the “technology gap” in an industry and the effect competition has on innovation. When the gap is relatively
small, competition tends to have a positive impact on innovation. When the gap is intermediate, an inverted-U arises. And when the gap is large, competition tends to have a negative impact on innovation. Thus, he argues that the difference in empirical results stems from a difference in the data-specific average technological gaps.

In this chapter, I estimate a model similar in spirit to the non-linear models above. The key differences are the following. One, I use a measure of competition that is consistent with the definition of “more competitive” established in Chapter 2. It is estimated as the elasticity of firm market value with respect to sales. Assuming sales are proportional to consumer value and market value mirrors consumer performance, my measure of competition captures the hypothetical number of sales a firm could have lost if it generated less consumer value than it otherwise did. Two, I use productivity as my measure of innovation. Non-linear models have principally used citation-weighted patent counts as a measure of innovation under the assumption that citations move monotonically with economic value. However, recent evidence suggests an inverted-U relationship. Three, I estimate the competition-innovation relationship at the firm level. While my measure of competition is an industry level measure, I allow productivity to vary at the firm level. The substantial increase in observations naturally assists in identification of the empirical parameters and provides a stronger result. Fourth, and finally, I allow for dynamic productivity growth. Dynamic productivity operates under the assumption that productivity-related decisions are based on past productivity and that output can take time to adjust to its new long-run level if the factors of production change (Nickell 1996).

The rest of the chapter is organized as follows. Section 3.1 gives a description of the empirical model used to estimate the competition-innovation relationship; discusses the standard methods used to estimate dynamic panel models; and finally talks
about what variables are assumed to be predetermined. Section 3.2 addresses the empirical measurement of innovation. Section 3.3 addresses the empirical measurement of competition. Section 3.4 gives an overview of the data. Section 3.5 discusses the empirical results. Section 3.6 discusses robustness checks. Section 3.7 concludes. And finally, section 3.8 offers some areas for future work.

3.2 The Model

The model takes the following form:

\[
y_{it} = \sum_{L=1}^{3} \left( \rho_{L}^y y_{it-L} + \rho_{L}^s s_{it-L} + \rho_{L}^m m_{it-L} \right) + \beta_1 c_{jt-1} + \beta_2 c_{jt-1}^2 \\
+ \rho_0^m m_{it} + \beta_3 n_{jk-t-1} + \beta_4 r_{jt} + d_t + tD_{jk} + \nu_{it} \tag{3.1}
\]

\[
\nu_{it} = \eta_i + \epsilon_{it}
\]

where, for firm \(i\) in industry \(j\), sub-industry \(k\) and during period \(t\), \(y\) is log multifactor productivity; \(s\) is scale, defined as log market value; \(c\) is competition; \(n\) is the number of sub-industry firms; \(m\) is a merger/acquisition dummy variable; \(r\) is a cyclical component, defined as the median industry return; \(d\) represents year fixed effects; \(D\) represents sub-industry fixed effects; \(\eta\) is unobserved firm heterogeneity, which is correlated with the productivity lags by construction; and \(\epsilon\) is an idiosyncratic error term. The parameters of interest are \(\beta_1\) and \(\beta_2\).

There are two classical variables that, from a competitive pressure perspective, may impact firm behavior to innovate: firm size and number of firms. \footnote{I include both.} There is no Herfindahl-Hirschman index for reasons given in Chapter 2. That said, I did explore its inclusion and found an economically and statistically insignificant relationship to productivity.
Regarding firm size, I use market value as a proxy as opposed to, say, employees or assets, because it not only accounts for technical size, but also current and expected future success. Along the lines of the so-called “quiet life hypothesis” proposed by Hicks (1935)—which essentially posits an inverse relationship between market power and efficiency—it is possible that current success contribute to “complacent” firm behavior in the future. Regarding the number of firms, this variable gives some sense of market structure and entry and exit activity. It is possible, for example, that firms alter their behavior in response to more competitors in the industry and enter and exit given their relative productivity.

Mergers and acquisitions are also likely to impact productivity. For example, synergies could aid a firm in achieving economies of scale in production or distribution or perform functions like research and development more efficiently (Farrell and Shapiro (2001)). On the other end, it is also possible for a merger or acquisition to stifle the innovation process, particularly if there are overlapping technologies that limit the scope of newer discoveries (Cassiman et al. (2005)). Whatever the case, there is reason to believe that mergers and acquisitions have some impact on the innovation process. Regarding how this is incorporated into the model, I include current as well as lagged merger and acquisition dummies. This is done for two reasons. One, I want to account for the possibility that mergers/acquisitions have an immediate and delayed impact on innovation. And two, by controlling for past merger and acquisition activity, we get a more accurate sense of actual industry exit in the event the number of firms decreases.

Past productivity is also likely to have an impact on future productivity investment

---

The number of publicly traded firms is obviously an imperfect measure of market structure because private firms are not accounted for.
decisions. By controlling for past productivity, we not only account for persistence and another dimension of heterogeneity (which helps in attaining consistent estimates of the other parameters), we also aid in eliminating serial correlation in the error term. The inclusion of lagged productivity is also consistent with the assumption in Olley and Pakes (1996) that productivity evolves dynamically, where they assume a first-order Markov process.

Finally, this model controls for macroeconomic shocks via time dummies; firm-specific, unobserved features that are time-invariant but possibly contribute to productivity (e.g., a company culture focused on innovation, a company that uses its reputation to consistently recruit and retain the “top minds,” a company adept at identifying and exploiting new technological trends, etc.); industry cyclicality via the time variation in median industry returns; and sub-industry specific trends.

With the exception of the macroeconomic and merger and acquisition variables, note that all covariates are lagged one period. This specification was chosen under the assumption that changes in firm behavior, specifically as it relates to future productivity choices, manifest only after the state of the environment has been realized. That is, companies take some time to adapt to their “new” environment. Take market value, for example. The simultaneous correlation between market value and productivity is not informative because market value is driven by performance, and performance is in large part driven by how productive a company is. By instead examining the relationship between lagged market value, say one period, and productivity, we ask a more interesting question: how does a company respond to its recent evaluation by the market? If the market lowers its evaluation of a company, perhaps said company will take it as a signal to improve its performance. On the other end,

\[\text{I use their method to estimate productivity.}\]
maybe a glowing evaluation by the market will make the company feel comfortable and give it less incentive to improve its performance. This same logic applies to the competitive state of the market and market structure. The objective is to ultimately see how productivity decisions respond to the state, not how they move with it.

Nickell (1996) estimates a similar model to 3.1. The main differences between my approach and Nickell’s are the following. One, I do not restrict the sum of the input elasticities to one. Two, I estimate the input elasticities industry-by-industry. And Three, I estimate the competition-innovation relationship in two stages as opposed to one. Regarding the last difference, I first estimate competition and the production function industry-by-industry, then I estimate the effect of competition on innovation using 3.1. In contrast, Nickell estimates the relationship in one stage, whereby value-added is made a direct function of the competitive environment and position of the firm.

The next section details estimation of the canonical dynamic panel model (See Roodman (2006) for an excellent survey on dynamic panel model estimation, which I follow here.)

Nickell estimates:

\[ y_{it} = \beta_i + \beta_t + \lambda y_{it-1} + (1 - \lambda)\alpha_i n_{it} + (1 - \lambda)(1 - \alpha_i)k_{it} + \alpha h_{it} + c_{it} + c_i t + \epsilon_{it}, \]

“where \( y_{it} \) is log real (value-added) output, \( n_{it} \) is log employment, \( k_{it} \) is log capital stock, \( h_{it} \) is a cyclical component, \( c_{it} \) reflects all factors capturing the impact of competition on the level of productivity, \( c_i \) reflects those factors that cover the impact of competition on productivity growth and \( [\alpha_i \) is employment elasticity for firm \( i \) in period \( t \) ... and \( \beta_i \) and \( \beta_t \) are firm effects and time effects.”
3.2.1 Estimation of Dynamic Panel Models

For notational simplicity, consider the canonical dynamic panel model:

\[ y_{it} = \rho y_{it-1} + x_{it}' \beta + \eta_i + \epsilon_{it}, \]

where \( x \) is assumed to be strictly exogenous. By construction, consistent estimation of the parameters is not possible via ordinary least squares. This is true since

\[ \mathbb{E}[y_{i,t-1} \eta_i] = \mathbb{E}\left[ (\rho y_{it-2} + x_{it-1}' \beta + \eta_i + \epsilon_{i,t-1}) \eta_i \right] \neq 0, \]

which follows from the fact that at least \( \mathbb{E}[\eta_i^2] \neq 0 \). Nevertheless, we can remove this endogeneity by transforming the model. Two commonly used transformations are first-differences and forward orthogonal deviations.

Transformations

The first-difference transformation subtracts the lag of the level equation from its current level, i.e.,

\[ \Delta y_{it} = \rho \Delta y_{it-1} + \Delta x_{it}' \beta + \Delta \epsilon_{it}, \quad (3.2) \]

and the forward orthogonal deviations transformation subtracts the forward average, over time, of the level equation from its current level, i.e.,

\[ y_{it+1}^* = \rho y_{it}^* + x_{it+1}' \beta + \epsilon_{it+1}^*, \]

where, in general, \( z_{i,t+1}^* = c_{it} \left( z_{it} - \frac{1}{T_{it}} \sum_{s>t} z_{is} \right) \), \( T_{it} \) is the number of observations for firm \( i \) after period \( t \) and \( c_{it} = \sqrt{\frac{T_{it}}{T_{it+1}}} \) is a scaling factor. The main appeal of the less commonly used forward orthogonal deviations transformation is that it retains more information than first-differencing in the presence of time gaps. For example, if \( y_{it} \) is missing, then a first-difference transformation will lead to two missing observations.
for firm $i$: $\Delta y_{it}$ and $\Delta y_{i,t+1}$. The orthogonal deviations transformation, on the other hand, will only lose $y^*_{i,t+1}$.

Regardless of the transformation used, the transformed lag of the dependent variable and the transformed idiosyncratic error term become correlated. In the case of the first-difference transformation, the variable $\Delta y_{i,t-1}$ is correlated with $\Delta \epsilon_{it}$ via the correlation between $y_{i,t-1}$ and $\epsilon_{i,t-1}$. And in the case of the forward orthogonal deviations transformation, the variable $y^*_{it}$ is correlated with $\epsilon^*_{i,t+1}$ via the correlation between $(y_{it}, y_{i,t+1}, ..., y_{iT})$ and $(\epsilon_{it}, \epsilon_{i,t+1}, ..., \epsilon_{iT})$. To purge the endogeneity, lags of the dependent variable beyond $t-1$ ($t$) may be used as instruments under a first-differences (orthogonal deviations) transformation. This is discussed further in the next section.

**Estimation of the Transformed Equation**

To purge endogeneity from the transformed equation, it is standard to use lagged levels and/or differences of the endogenous regressors as instruments. Anderson and Hsiao (1982) were the first to propose this basic strategy, namely a 2SLS estimator that uses either $y_{i,t-2}$ or $(y_{i,t-2} - y_{i,t-3})$ as an instrument. Holtz-Eakin et al. (1988) and Arellano and Bond (1991) point out, however, that while the method proposed by Anderson and Hsiao results in consistent estimates, it is not the most efficient. This is because it does not leverage the information contained in all the lags of the endogenous regressors, and because the instrument matrix drops observations where there are lags. To avoid these shortcomings, Holtz-Eakin et al. (1988) developed an observation-preserving system (i.e., a system of equations, one per time period) GMM estimator that utilizes all suitable lags of the endogenous regressors in levels as instruments. Known as the “difference” GMM estimator, it is widely used in dynamic
panel settings. Its moment conditions correspond to \(3.2\) and are given by

\[
\mathbb{E}[y_{i,t-2}\Delta \epsilon_{it}] = 0 \quad \text{for each } t \geq 3.
\]

Another commonly used estimator, the so-called “system” GMM estimator, uses additional moment conditions linked to the levels equation \(3.1\). Proposed by Arellano and Bover (1995) and Blundell and Bond (1998), the system GMM estimator was developed because levels may be weak predictors of differences. That is, difference GMM may suffer from weak instruments.\(^5\) As a result, Blundell and Bond developed an estimator that estimates the stack of \(3.1\) and \(3.2\) using

\[
\mathbb{E}[\Delta y_{i,t-1}\nu_{i,t}] = 0 \quad \text{for each } t \geq 2.
\]
as additional moment conditions.

Both estimators were developed with a small time dimension, \(T\), in mind. In fact, the number of instruments implied by the moment conditions are quadratic in \(T\). Large \(T\) panels may therefore suffer from biased standard errors, overfitting of the endogenous variables and/or weakened Hansen J tests.\(^6\) Fortunately, there are several ways to deal with this. One way is to simply use less than \(T\) lags. Another

\(^5\)In fact, Blundell and Bond demonstrate that if a variable’s time-series is close to a random walk, then difference GMM performs poorly. To see this, consider

\[
y_{it} = \alpha + \rho y_{i,t-1} + x_{it}'\beta + \epsilon_{it}.
\]

We can rewrite this as

\[
\Delta y_{it} = \alpha + (\rho - 1)y_{i,t-1} + x_{it}'\beta + \epsilon_{it}.
\]

If \(\rho\) is close to 1 (i.e., \(y_{it}\) is close to a random walk), the correlation between \(\Delta y_{it}\) and \(y_{i,t-1}\) will be close to zero, implying that the levels of \(y\) are weak instruments for differences in \(y\).

\(^6\)See Roodman (2009)
way is to operate with less moment conditions. In the case of difference GMM, for example, we may use the following moment conditions as an alternative:

\[ \sum_i \sum_t y_{i,t-2} \Delta \epsilon_{it} \rightarrow \mathbb{E}[y_{i,t-2} \Delta \epsilon_{it}] = 0, \]

which makes the number of instruments linear in \( T \). And the final way would be some combination of using less lags and collapsing the instrument set via the former.

It is not possible to discern, \textit{a priori}, which estimator and moment conditions are appropriate. We have at our disposal, however, several tests, such as the Hansen J and Arellano-Bond tests, to help guide us. I explore various methods and find that difference GMM with a collapsed instrument set works best (in terms of instrument validity, serial correlation of the idiosyncratic error term and a theoretically sensible estimate of the first-order autoregressive parameter).

**Post-estimation Tests of Instrument Validity**

Like any estimator that uses instruments to achieve consistency, validity of the instruments should be tested. Two tests are standard in a dynamic setting: (1) the Hansen J test for overidentifying restrictions and (2) the Arellano-Bond test for serial correlation of \( \epsilon_{it} \). The Hansen J test tests the validity of instruments under the assumption that the instruments are valid, or uncorrelated with the residuals. Thus, if the cross-product between the instruments and the residuals is “too large,” the test will reject the null.

The Arellano-Bond test, on the other hand, tests the validity of instruments under the assumption of a non-serially correlation \( \epsilon_{it} \). If the idiosyncratic error term \( \epsilon_{it} \) is serially correlated, then there is a possibility of correlation between the lags of the dependent variable and \( \epsilon_{it} \). For example, if \( \epsilon_{it} \) is first-order serially correlated, \( y_{it-2} \)
will be correlated with $\Delta \epsilon_{it}$, casting doubt on $y_{it-2}$ as a valid instrument. In practice, the Arellano-Bond test for serial correlation is applied to the first-differenced residuals, $\Delta \hat{\epsilon}_{it}$ (since we are estimating [3.2]). The Arellano-Bond test thus tests for first-order serial correlation of $\epsilon_{it}$ under the assumption that second-order serial correlation of $\Delta \epsilon_{it}$ is zero.\(^7\)

I use both the Hansen J and Arellano-Bond tests to guide model selection. If, for example, one model strongly rejects the null of the Hansen J and Arellano-Bond tests, while another strongly fails to reject both, one can infer the the former is misspecified.

In addition to the above tests, I compare the results of each model to the results that are generated by the fixed-effects and ordinary least squares estimators. It is well known that the fixed-effects estimator underestimates the first-order autoregressive coefficient, and the ordinary least squares estimator overestimates it. We are therefore given a rough bound on where said coefficient should lie. If the coefficient lies outside the bound, then there is reason to cast doubt on the specification of the model.

### 3.2.2 Choice of Instruments

The model given by [3.1] assumes more than just one predetermined regressor. In addition to the productivity lags, it is reasonable to assume that shocks to productivity have an impact on market value and the number of firms in the industry (See Syverson (2004) for an example of how productivity can affect the number of firms in an industry). For example, a positive shock to productivity is likely to have a positive impact on a firm’s market valuation, and a negative shock to productivity could

\(^7\)First-order serial correlation between $\Delta \epsilon_{it}$ and $\Delta \epsilon_{it-1}$ will exist by construction. However, there need not be serial correlation between $\Delta \epsilon_{it}$ and $\Delta \epsilon_{it-2}$. If there is, this suggests a correlation between $y_{it-1}$ and $y_{it-2}$.
induce relatively unproductive firms to exit their market. Thus, the first-difference transformation will induce a correlation between the one-period lag of the idiosyncratic error term, $\epsilon_{it-1}$, and the one-period lags of market value, $s_{it-1}$, and number of firms, $n_{jk,t-1}$. I consequently use lags of market value and number of firms as instruments to assist in mitigating the effect of endogeneity.

Regarding competition, the literature has taken the approach that innovation and competition are to be treated as endogenous. This is logical given the recent trend toward profit margins as a measure of competition. For example, process innovation will, all else equal, increase a firm’s profit margin by lowering its costs. My measure of competition, however, is not defined in terms of technical market power; instead, it captures how much market value a firm would have lost in expectation if it generated less value than it otherwise did. This amounts to an estimate of market value elasticity with respect to sales. Shocks to productivity are therefore unlikely to have a meaningful impact on my measure of competition. It is particularly unlikely in my model given that innovation and competition are measured at the firm and industry level, respectively. That is, it is reasonable to assume that shocks to productivity at the firm level have effectively no impact on the expected loss in market value that stems from lower sales.

Regarding the number of exclusions restrictions used for each predetermined variable, I use two in my baseline model. However, I explore deeper lags of the predetermined variables as instruments as a robustness check.
3.3 Measures

3.3.1 Measure of Innovation

As discussed in Chapter 2, there are three common measures of innovation in the empirical literature: R&D expenditures, patents and productivity. I use productivity and summarize why below:

1. R&D expenditures are an input to, not an output of, the innovation process. Moreover, it is plausible that even as a proxy of effort a decrease in R&D expenditures could coincide with an increase in innovative output. For example, if a firm switches from exploratory to exploitative R&D, innovative output could potentially increase despite a decrease in expenditures.

2. Patent statistics, such as raw and citation-weighted patent counts, are compromised on several fronts. One, the economic incentive to apply for a patent is not confined to protecting valuable intellectual assets (e.g., a company might apply for a patent with the intent of protecting itself from possible litigation, not because the patent itself embodies a novel and valuable idea). Two, patents represent a small fraction of innovation. And three, specifically as it pertains to citation-weighted patent counts, evidence indicates that the presumption of monotonicity between value and number of citations is invalid.

3. Productivity captures innovation from all sources, not just those that are patented. It is also intuitive. For example, if a company makes a substantial and successful investment toward the reduction of its production costs, then a rise in productivity will follow, all else equal. Likewise, if a company introduces an innovative product to the market, then a rise in productivity will follow (the extent of this rise will,
however, depend on factors like brand loyalty, switching costs, etc.). Finally, and most importantly, productivity growth coincides with long-run economic growth.

Productivity estimation is itself a topic of substantial interest in the economics literature (See Beveren (2007) for a concise survey of the literature, which I follow). To motivate, consider the following production function:

$$ Y_{it} = A_{it} L_{it}^{\beta_L} K_{it}^{\beta_K}, $$

where, for firm $i$ in period $t$, $Y$ is output/value-added, $L$ and $K$ are labor and capital inputs, respectively, and $A$ is the Hicksian-neutral level of efficiency. Taking logs, we have

$$ y_{it} = \beta_0 + \beta_L l_{it} + \beta_K k_{it} + \epsilon_{it}, $$

(3.3)

where lower case variables are denote logs and $\ln(A_{it}) = \beta_0 + \epsilon_{it}$. The variable $\epsilon$ is a random variable commonly known as the Solow residual; it captures productivity, or output per unit of input. To compute productivity, the equation above is estimated via regression.

It was once standard to estimate (3.3) via OLS using a balanced panel of firms. However, OLS suffers from two sources of bias. One is selection bias. The selection bias emerges from the fact that, despite a balanced panel, exit and entry decisions are influenced by productivity differences. Farinas and Ruano (2005), for example, find that Spanish manufacturing firms are systematically more likely to exit their industry if their level of productivity is significantly below the frontier. The other source of bias stems from simultaneity between input choices and current productivity. For example, a firm is likely to use more labor than it otherwise would if its unobserved level of productivity is high.
Several methods have been proposed to deal with selection and simultaneity bias. Instrumental variables and generalized method of moments estimators are one approach. However, as is always the case with instrumental variable estimators, valid instruments are required to achieve consistent estimates of the parameters. In practice, such instruments are difficult to find. Consider prices. One might opt to use factor prices as instruments for factor quantities, but if factor markets are not perfectly competitive, shocks to output will impact factor prices. Another proposal is fixed effects estimation. But this method only works if it is reasonable to assume that unobserved productivity remains constant over time, which is unlikely. Finally, we have semiparametric estimators. Semiparametric estimators are computationally more demanding than fixed effects and instrumental variable estimators, but for the purposes of estimating productivity, they are superior; they explicitly account for selection bias, and they account for unobserved productivity via an equilibrium condition that allows one to express it in terms of observable variables.

Two of the most widely used semiparametric estimators are Olley and Pakes (1996), or OP, and Levinsohn and Petrin (2003), or LP. These estimators differ in the way they control for unobservable productivity. OP expresses unobservable productivity as a function of investment, whereas LP expresses it in terms of intermediate inputs. The choice of which estimator to use depends in part on the data, but also on a complication that arises from collinearity.

The OP method requires monotonicity between investment and unobserved productivity. In terms of what this means practically, if data on investment is “lumpy,” then estimates of the input elasticities will be inefficient. This is precisely why LP introduced their method: to avoid the issue of zero investment in the data. This potential advantage of LP is, however, marred by the possibility of unidentifiability of
the labor coefficient. Ackerberg et al. (2015) show that if labor and materials (the intermediate input choice of LP) are simultaneously chosen together, then identification of the labor coefficient is not possible.

For my purposes, I use OP to estimate productivity. Their model now follows.

**Productivity Estimation**

To estimate productivity, I follow Olley and Pakes (1996). The advantage of OP’s estimation procedure is that it accounts for selection and simultaneity bias. The selection bias emanates from the relationship between the firm’s unobserved (to the econometrician) productivity and its exit decision, and the simultaneity bias emanates from the relationship between productivity and input demands.

Before OP introduced their algorithm, it was traditional to construct a balanced panel of firms to control for entry and exit and then apply OLS or the “within” estimator to estimate the production function consistently. The disadvantages of this approach include removal of potentially a substantial amount of observations and the assumption that unobserved productivity is fixed over time. OP obviates the need to construct a balanced panel and controls for simultaneity bias via a semiparametric estimation procedure. Their model is now outlined.

**Olley and Pakes Estimation Procedure**

The OP model begins with the behavior of the firm. Firms are risk neutral and maximize their expected discounted value of future net cash flows. Current profit, \( \pi_t(\cdot) \), for any firm depends on a set of firm-specific state variables, factor prices and a vector that lists the state variables of the other firms. The state variables are a firm’s age, \( a_t \), capital stock, \( k_t \), and efficiency, \( \omega_t \).
Every period, a firm faces potentially three decisions. At the beginning of the period, the firm must decide whether or not to exit the industry. If the scrap value, $\Sigma$, today is greater than its expected future value, it will exit today and never reappear again. Otherwise, the firm will remain in the industry and decide on how much labor to employ, $l_t$, and how much to invest toward new capital, $i_t$. The transition equations for capital and age are:

\[
\begin{align*}
  k_{t+1} &= (1 - \delta)k_t + i_t \\
  a_{t+1} &= a_t + 1,
\end{align*}
\]

where $\delta$ is the depreciation rate of capital. Productivity evolves according to an exogenous Markov process that is known by the firm.

The Bellman equation that the firm faces is given by

\[
V_t(a_t, k_t, \omega_t) = \max \left\{ \Sigma, \sup_{i_t \geq 0} \pi_t(a_t, k_t, \omega_t) - c(i_t) + \beta \mathbb{E}[V_{t+1}(a_{t+1}, k_{t+1}, \omega_{t+1}) | I_t] \right\},
\]

subject to the transition equations for age, capital and productivity. The parameter $\beta$ is the discount factor, $c(\cdot)$ is the cost of investment, $\mathbb{E}[\cdot]$ is the expectation operator, and $I_t$ is information at time $t$. OP show that there exists a Markov perfect Nash equilibrium to this problem. In addition to an optimal investment decision, the solution to this problem generates an exit rule. The exit rule, $\chi_t$, and optimal investment decision take the following general forms, respectively:

\[
\chi_t = \begin{cases} 
1 & \text{if } \omega_t \geq \omega_t(a_t, k_t) \\
0 & \text{otherwise}
\end{cases}
\]

and

\[
i_t = \iota_t(a_t, k_t, \omega_t),
\]

(3.6)
where \( \omega(t) \) and \( i_t(t) \) are determined in equilibrium. The exit rule is intuitive: a firm will remain in the industry if and only if its productivity level is at least as large as some threshold value, \( \omega \), generated by the model. These two pieces of the model are essential for removing the selection and simultaneity biases mentioned above. Specifically, the exit rule takes care of the selection bias, and the optimal investment function takes care of the simultaneity bias. Putting more structure on the production function, as we are now about to do, will make this more clear.

Following OP, assume the industry offers a homogeneous product using a Cobb-Douglas production technology. That is, suppose

\[
y_{it} = \beta_0 + \beta_l l_{it} + \beta_k k_{it} + \beta_a a_{it} + \epsilon_{it} \\
\epsilon_{it} = \omega_{it} + \eta_{it},
\]

(3.7)

where \( y_{it} \) is log output (value-added), \( l_{it} \) is log labor, \( k_{it} \) is log capital, \( a_{it} \) is age, \( \omega_{it} \) is productivity and \( \eta_{it} \) is an i.i.d. zero-mean disturbance for firm \( i \) at time \( t \). As discussed above, the simultaneity bias arises because the firm’s capital and labor choices are correlated with its current productivity, i.e., \( \mathbb{E}[\omega_{it} | l_{it}, k_{it}, a_{it}, i_{it}] \neq 0 \).

In theory, we could find instruments for capital, labor and age and estimate using standard IV methods, but finding valid instruments is hardly convincing. OP instead apply the predictions of their behavioral model to achieve consistent estimates of the parameters.

The procedure goes as follows. We first observe that the optimal investment decision (see 3.6) of the firm is a function of its age, capital stock and productivity. OP show that if \( i_t > 0 \), then the firm’s optimal investment decision is strictly increasing in \( \omega_t \) for any pair \( (a_t, k_t) \). Consequently, we may invert \( i_t \) to rewrite the firm’s unobserved
productivity as a function of its investment, age and capital stock. That is,

\[ \omega_t = h_t(i_t, a_t, k_t), \quad (3.8) \]

where \( h_t = i_t^{-1} \). Substituting this into \( 3.7 \) gives

\[ y_{it} = \beta_l l_{it} + H_t(i_{it}, a_{it}, k_{it}) + \eta_{it}, \quad (3.9) \]

where \( H_t(i_{it}, a_{it}, k_{it}) \equiv \beta_0 + \beta_k k_{it} + \beta_a a_{it} + h_t(i_{it}, a_{it}, k_{it}) \). By implementing a second-order (or higher) polynomial in investment, age and capital stock to approximate \( H_t(\cdot) \), we control for unobserved productivity, and this allows us to obtain a consistent estimate of \( \beta_l \) in \( 3.9 \) via OLS. Performing OLS on \( 3.9 \) however, does not identify \( \beta_k \) and \( \beta_a \), because \( 3.9 \) does not separate the effect of capital and age on investment (see \( 3.6 \)) from their effect on output. Nevertheless, we can isolate the simultaneous effect of capital and age on output by leveraging information from the previous period. To see this, rewrite \( 3.7 \) as

\[ y_t - \beta_l l_t = \beta_0 + \beta_k k_t + \beta_a a_t + \omega_t + \eta_t \]

and take the expectation conditional on information at time \( t \) and survival, giving

\[ \mathbb{E}[y_t - \beta_l l_t \mid I_t, \chi_t = 1] = \beta_0 + \beta_k k_t + \beta_a a_t + \mathbb{E}[\omega_t \mid \omega_{t-1}, \chi_t = 1] \]

\[ = \beta_k k_t + \beta_a a_t + g(\omega_t, \omega_{t-1}). \quad (3.10) \]

The function \( g(\omega_t, \omega_{t-1}) \equiv \beta_0 + \mathbb{E}[\omega_t \mid \omega_{t-1}, \chi_t = 1] \) is the key to identifying \( \beta_k \) and \( \beta_a \). Recall that capital and age have a simultaneous effect on output and investment and this is why we cannot identify \( \beta_k \) and \( \beta_a \) using \( 3.9 \) as-is. However, we can express \( i_t \) in terms of \( i_{t-1} \) and thereby circumvent this issue. Observe that survival at time \( t \)
depends on information at time \( t - 1 \), i.e., investment, capital and age at time \( t - 1 \):

\[
\mathbb{P}(\chi_t = 1|\omega_t, I_{t-1}) = \mathbb{P}(\omega_t \geq \omega_t(a_t, k_t)|\omega_t(a_t, k_t), \omega_{t-1})
\]

\[
= \mathbb{P}(\omega_t(a_t, k_t), \omega_{t-1})
\]

\[
= \mathbb{P}(a_{t-1}, k_{t-1}, i_{t-1})
\]

\[
\equiv \mathcal{P}_{t-1}. \tag{3.12}
\]

The third line follows from the fact that \( \omega_{t-1} \) is direct function of information at time \( t - 1 \), and \( \omega_t \) is an indirect function of information at time \( t - 1 \) (the transition equations given by (3.4) imply that \( k_t \) and \( a_t \) are functions of \( k_{t-1} \) and \( a_{t-1} \), respectively).

Therefore, since \( g(\omega_t, \omega_{t-1}) \) may be expressed as a function solely of information at time \( t - 1 \), we can isolate and identify the effect of capital and age on output.

The only obstacle to carry out the above is that we need a measure of \( \omega_t \) and \( \omega_{t-1} \).

To measure the latter is straightforward. Observe that \( \omega_{t-1} = H_{t-1} - \beta_k k_{t-1} - \beta_a a_{t-1} \).

Thus, we need a consistent estimate of \( H_{t-1} \) before we can estimate \( \omega_{t-1} \), which we obtain by estimating (3.9). Specifically, \( \hat{H}_{t-1} = \hat{y}_{t-1} - \hat{\beta}_h l_{t-1} \). To estimate \( \omega_t \), observe that it can be expressed as a function of \( \mathcal{P}_{t-1} \) and \( \omega_{t-1} \) (see (3.12), i.e., we may rewrite \( g(\omega_t, \omega_{t-1}) \) as \( g(\mathcal{P}_{t-1}, \omega_{t-1}) \). This indicates that we need an estimate of \( \mathcal{P}_{t-1} \), which we obtain by a Probit regression of \( \chi_t \) on some function of \( i_{t-1}, a_{t-1} \) and \( k_{t-1} \).

Ultimately, we are left performing non-linear least squares on

\[
y_t - \hat{\beta}_l l_t = \beta_k k_t + \beta_a a_t + g(\hat{\mathcal{P}}_{t-1}, \hat{H}_{t-1} - \beta_k k_{t-1} - \beta_a a_{t-1}) + \xi_t + \eta_t \tag{3.13}
\]

to identify \( \beta_k \) and \( \beta_a \), where \( \xi_t \equiv \omega_t - \mathbb{E}[\omega_t | \omega_{t-1}, \chi_t = 1] \) and \( g(\cdot) \) is approximated with a second- or higher-order polynomial. This concludes the OP algorithm.

For each 3-digit NAICS industry\(^8\) I estimate a model similar to OP, but with two adjustments given the structure of my data. One, to account for heterogeneity

\(^8\)The 3-digit level was chosen because it offers a relatively high number of observations, on
across more narrowly defined industries, I include sub-industry dummies at the 6-digit NAICS level. And two, I include a quadratic, as opposed to a linear, time trend to crudely account for any partial non-linearities in productivity over time. Once the model is estimated, log productivity is calculated as $\ln(\text{TFP}) \equiv y - \hat{\beta}_l l - \hat{\beta}_k k$.

3.3.2 Measure of Competition

To measure competition in a way that is consistent with the contestability principle proposed by Shapiro (2012), I adopt a method similar to Boone et al. (2005). Boone’s view is that a more competitive industry is identified by how harshly relatively inefficient firms are punished. Specifically, he measures the degree of competition in an industry by how small (large) the profits are of relatively inefficient (efficient) firms. This, in turn, is approximated by the slope coefficient corresponding to the regression of (log) gross profits onto (log) average variable costs. There are two issues with this approach, however. One, gross profits and variable costs are mathematically related by construction, and two it assumes that efficiency is the only way for a firm to advance or protect its competitive position. The latter is self-explanatory, and the former is problematic because the coefficient of interest is uninterpretable.9

average, across industries. This is particularly important given the finite sample properties of non-linear/semi-parametric estimators.

9While endogeneity in and of itself is not problematic given a set of valid instruments, the form that it takes in this particular case is not rectifiable. To see this, note that Boone defines average variable costs as the ratio of costs of goods sold to revenue. Let $\pi$, $s$, $c$ and $x \equiv c/s$ denote gross profit, revenue, cost of goods sold and average variable costs, respectively. Then gross profit may be expressed as $\pi = s - c = s(1 - x)$, making the marginal effect of average variable costs on gross profits uninterpretable in a regression setting (even if, while highly unlikely, a set of valid instruments existed).
For the above reasons, I propose an alternative but similar method that captures the expected loss (gain) in firm market value (assigned by the stock market) that would hypothetically occur if a firm had otherwise generated less (greater) sales. I approximate this by the elasticity of market value with respect to sales. This estimate is consistent with the contestability principle if sales are indicative of (and signal as much to the market) consumer value/success—be it due to superior products, services, business models, etc.—and if the value of a firm assigned by the market is based on its relative performance. Assuming this holds, if the expected loss (gain) in market value increases from one year to the next, then the subsequent year may be viewed as more contestable.

For each industry $j$ at the 3-digit NAICS level, the degree of contestability at time $t$ is estimated via the following:

$$\ln (V_{it}) = \sum_t \phi_{jt} d_t \ln (S_{it}) + \beta \ln (C_{it}) + \gamma m_{it} + t D_{jk} + d_t + \alpha_i + \epsilon_{it}, \quad j = 1, \ldots, J,$$

where, for firm $i$ in industry $j$, sub-industry $k$ at time $t$, $V$ is market value; $S$ is sales; $C$ is operating expenses; $m$ is a dummy variable that captures a merger or acquisition; $d$ are a set of time dummies; $D$ are a set of sub-industry dummies at the 6-digit level; $\alpha$ is unobserved, time-invariant firm heterogeneity; and $\epsilon$ is an idiosyncratic error term. The parameters of interest are $\phi = (\phi_{11}, \phi_{12}, \ldots, \phi_{jt}, \ldots, \phi_{JT})$.

Ideally, estimation of the above equation would be done with a more narrow industry classification; however, this greatly limits the ability to identify the set of $\phi$ parameters.\footnote{In other words, sales may be treated as an index of consumer-generated value.\footnote{Intuitively, each $\phi$ is a cross-sectional estimate. Thus, by narrowing the industry classification, we lose degrees of freedom for each year in the sample.}} In fact, doing so results in many of the $\phi$ parameters being negative.
or imprecisely estimated. My strategy is thus to estimate the degree of contestability with a more broad industry classification, the 3-digit NAICS level, but include sub-industry dummies at the 6-digit NAICS level.

The market value of a firm and its evolution over time depends on several factors outside sales. For example, macroeconomic shocks, like the Great Recession, and the business cycle will have an effect on trading volume and the propensity to buy or sell. Likewise, idiosyncratic features, such as cost structure, mergers and acquisitions and various time-invariant, unobserved characteristics (e.g., product scope), will affect the variability of a firm’s market value. I account for these factors with the inclusion of time dummies, a variable that controls for the firms cost structure ($C$), a dummy variable that captures the event of a merger or acquisition ($m$) and an unobserved, firm-specific factor ($\alpha$) that is assumed to be correlated with both sales and costs.

Given the assumption that the unobserved, firm-specific factors are correlated with sales and costs, ordinary least squares performed on Equation 3.3.2 will result in inconsistent estimates of the parameters. I therefore estimate 3.3.2 with the fixed effects estimator.

The competition variable, $c$, used in Equation 3.1 is defined as

$$c_{jt-1} = \frac{\hat{\phi}_{jt-1}}{\frac{1}{T_j} \sum_{t=1}^{T_j} \hat{\phi}_{jt}},$$  \hspace{1cm} (3.14)

where $T_j$ is the number of yearly observations for industry $j$, and $\hat{\phi}$ is the estimated value of $\phi$. Thus, I normalize the $\phi$ parameters by their industry means over time. This is done to mitigate any noise due to industry heterogeneity and to help with

---

12 One additional step is taken before implementing the measure of competition as defined. In particular, I center $c$ because of the high degree of correlation—approximately 0.99—between $c$ and $c^2$ in the estimation sample, and because the variance inflation factor is close to 60. After centering, the correlation drops to approximately -0.08, and the variance inflation factor drops to 1.
interpretation of what is clearly an abstract measure. Regarding the latter, a unit increase in competition now corresponds to $\phi$ increasing by an amount equal to its industry average.

In total, there are 893 industry-year estimates of $\phi$. Of these 893 estimates, 888 are positive. If negative estimates are ignored (the idea being that the market value of a firm should, on average, increase with sales), 16 firm-level observations will be lost—a negligible fraction of the sample (less than 0.1%). I consequently ignore these observations in the rest of my analysis.

3.4 The Data

Firm level data are sourced from Compustat, a database on publicly traded firms dating back to 1950. Annual data from 1950 to 2015 spans 463,608 firm-year observations across 36,089 firms. This data is merged with the NBER-CES Manufacturing Database, which has data on prices and quantities in the manufacturing sector. Price deflators at the 6-digit NAICS levels are extracted from the NBER-CES data set to construct measures of capital, labor, value-added and investment.

Before merging the Compustat and NBER productivity databases, some intermediate steps are taken. First, any firm in the Compustat sample for no greater than two years is dropped. Second, all firms in the Compustat sample without a NAICS code are kept and merged using an SIC to NAICS concordance table. Once this is

---

13 The 7 negative estimates all correspond to NAICS code 323, or the Printing and Related Support manufacturing industry.

14 The manufacturing sector consists industries that fall within 31 and 33 of the 2-digit NAICS classification.

15 Firms without a NAICS code are kept because all firms have an SIC code, which may then be converted to a NAICS code. The conversion from SIC to NAICS is facilitated by a concordance
done, all the firms in the Compustat sample with a missing NAICS code are merged with the concordance table. This leaves 137,228 firm-year observations and 8,601 firms intact.

After the intermediate steps are taken care of, the above is merged with the NBER-CES Manufacturing dataset. Doing so leaves 113,318 firm-year observations and 7,395 firms intact. Part of the reduction is due to the fact that some firms do not have 6-digit NAICS code. The other part of the reduction is simply due to non-matching data. I then drop all observations prior to 1962 following Fama and French (1992). They argue that before this year, the data are composed of a disproportionate number of large and highly successful firms.

The final sample before estimation has 59,251 firm-level observations from 1962-2009 with 4,749 firms and 126 6-digit industries. The large reduction in observations stems from firms that drop from the sample due to negative or missing values for value-added (i.e., sales minus materials is negative), capital, labor, investment and materials.

Estimation of firm-level productivity requires measures of labor, capital, investment and value-added. They are measured as follows (using 6-digit NAICS price deflators). Value added is net sales minus materials deflated by the price of output, where materials is calculated as operating expenses net of depreciation and labor expenditure deflated by the price of materials. Labor is the number of employees deflated by average wages. Investment is capital expenditures deflated by the price table provided by the U.S. Department of Commerce (see Census (2015)). The specific concordance table used is that which converts 1987 SIC codes to 1997 NAICS codes. Note that some SIC codes do not have a unique assignment to a 6-digit NAICS code. I remove these SIC codes, which reduces the number of manufacturing SIC codes from 472 to 310).
of investment. Capital is calculated using a perpetual inventory following Hall (1990) and Imrohoroglu and Tuzel (2014).

### 3.5 The Results

The main results are reported in table 3.1. For each estimator, namely the Holtz-Eakin, Newey and Rosen (HNR) (i.e., the “difference” GMM estimator), fixed effects (FE), ordinary least squares (OLS) and system GMM (SGMM) estimators, I estimate two models. Model (1) includes sub-industry dummies interacted with a time trend, and model (2) does not.

Observe first the Hansen J and Arellano-Bond (AB) $p$-values. Under the null of exogenous instruments, the HNR model with industry dummies (column (1)) indicates that there is an 85% chance of observing a statistic just as large as the one estimated. Likewise, under the null of no serial correlation, the same model indicates that there is a 59% chance of observing a statistic just as large as the one estimated. Compare this to the SGMM estimator with industry dummies. The SGMM estimator indicates that there is a 0% chance of observing a statistic just as large under the null of exogenous instruments, and only a 20% chance under the null of no serial correlation. For this reason, and the fact that the SGMM estimates closely resemble the OLS ones, I use the HNR estimates for the rest of my analysis.

The point estimates for past productivity are as expected for each estimator: positive and less than one in summation. Thus, productivity growth converges to a

---

16Given the size of the tables, table notes are not included. Note that the stars indicate levels of significance, namely $*p < 0.05$, $**p < 0.01$, $***p < 0.001$. The numbers in parentheses are $t$-statistics. Standard errors are robust to general forms of heteroskedasticity and autocorrelation at the firm level. Two-step GMM is performed using the Windmeijer correction.
steady state. Notice also that the first-order autoregressive coefficients under HNR fall between the FE and OLS ones. This accords with theory that OLS (FE) overestimates (underestimates) the autoregressive coefficient in a dynamic panel setting with unobserved heterogeneity. The autoregressive coefficient under SGMM likewise falls between the FE and OLS estimates, but it is much closer to the OLS point estimates. By itself, this is not too much cause for concern; however, the Hansen J test also strongly rejects exogeneity of the instruments, indicating that the model is misspecified.

The market value estimates are interesting in that both FE and OLS predict a positive impact from the one-period lag, while HNR predicts a negative impact. Moreover, the HNR estimates predict no statistically significant impact from the three-period lag, while FE and OLS (except FE without sub-industry dummies) predict otherwise. These differences are likely due to simultaneity. Each estimator, however, predicts similar and highly significant point estimates for the two-period lag. Thus, it appears size and success have a delayed negative impact on productivity.

On mergers and acquisitions, estimates are relatively similar across the board, both in terms of sign and significance. The sign is consistently the same for each period, with the largest and most statistically significant impact stemming from the two-period lag. This might suggest that mergers and acquisitions are initially costly (likely because it takes time to fully synergize), but afterwards, synergies are realized and manifest in terms of greater productivity.

The market structure estimate (i.e., number of firms) is equally strong across the board, with the HNR point estimates being the largest. In the hypothetical scenario of ten (one) firms entering the industry, the HNR model predicts an approximately 12% (1.2%) increase in productivity. Granted, this is a crude measure of market
structure, but it nevertheless has an economically significant impact on innovation.

Moving to competition, first note that the marginal effect of competition is given by

\[
\frac{\partial E[y | \cdot]}{\partial c} = \beta_1 + 2\beta_2 c
\]

Thus, the HNR results in column (1) imply the following estimated marginal effects

\[
\frac{\partial E[y | \cdot]}{\partial c} = 0.2 + 0.782c
\]

The positive estimate of \( \beta_2 \) implies that the marginal effect of competition on innovation is increasing. However, because \( c \) is centered, this does not rule out a negative marginal effect. For this reason, I plot the marginal effects over a discrete grid of \( c \) values that lie within its first and ninety-ninth percentiles of the estimation sample. See Figure 3.1.

From Figure 3.1, we immediately see two things: (1) the estimated marginal effect is positive for all values of \( c \) and (2) the 95% confidence interval (grey shaded area) includes non-positive values only for relatively low values of competition. In other words, while the marginal effect is positive everywhere, increases in competition have a statistically ambiguous effect on innovation when competition is relatively low. One possibility is the following. When competition is relatively low, the industry is in a state of stasis: innovation is both slow and mostly incremental. If firms therefore act on the belief that the industry is unlikely to significantly change given its current state, innovation will remain low. This accords with the quiet life hypothesis posited by Hicks (1935), which is that firms with market power (not in the technical sense defined by Lerner (1934)) are inclined not to innovate as this would generate supranormal rents and consequently the attention and uncertainty of rival entry.

While the marginal impact of competition is statistically unclear at relatively low
Figure 3.1: Marginal impact of competition on innovation: Positive $\phi$
levels of competition, this is not the case at relatively high levels. For values of $c > 1$, the marginal effect is unambiguously positive. Thus, it appears firms are more responsive to an increase in competition when competition is already high. Along the same lines as above, this could be the case if high levels of competition condition firms to believe that an increase in competition is credible; i.e., more weight is given to the risk of lost sales when an industry is already relatively competitive.

### 3.6 Robustness check

I perform two robustness checks. First, I analyze the sensitivity of the model with respect to changes in the lag structure of the instrument matrix. It is well understood that “too” many instruments can weaken the Hansen J test of overidentifying restrictions (Roodman (2006)). Intuitively, too many instruments can overfit the endogenous regressors or, because additional instruments are introduced via deeper lags, can introduce weak instruments, making it difficult to expunge endogeneity from the regressors.

The lag structure of my principal model uses a collapsed instrument set with two exclusion restrictions for each predetermined variable. For example, lagged productivity is instrumented with five lags, with the fourth and fifth being the two excluded lags. This lag structure was chosen primarily with the intent of mitigating the problem of weak instruments. It was also chosen for its theoretical appeal, as lags beyond three years are unlikely to be economically meaningful or statistically relevant. The results of this analysis can be seen in table 3.3. The column headers indicate how many exclusion lags were used for the predetermined regressors.

The results of the lag-sensitivity analysis indicate a negligible impact on the point estimates and standard errors. However, the $p$-value corresponding to the Hansen J
test monotonically decreases with lag count. This is somewhat counterintuitive given that deeper lags should, if anything, be orthogonal to the error term. Examination of the Hansen J test statistic may provide some intuition. It is computed as follows:

\[ J = \left( \frac{1}{NT} Z'E \right)' S^{-1} \left( \frac{1}{NT} Z'E \right), \]

where \( N \) is the number of panels, \( T \) is the number of time periods, \( Z \) is the instrument matrix, \( E \) is the residual matrix and \( S \) is the estimated covariance matrix of \( Z'E \).

Thus, the Hansen J statistic is the minimized value of the generalized method of moments criterion function, which is itself a quadratic function of the variance-normalized sample moments. Quick inspection might therefore suggest that inclusion of deeper lags should drive \( J \) to zero. However, if the inclusion of deeper lags as instruments introduces a weak instrument problem, inconsistent and inefficient estimates of the structural parameters may arise, the former implying that \( Z'E \) will not converge to zero. In any case, the Hansen J test indicates that a deeper lag structure is misspecified.

As another robustness check, I reestimate the model using relatively precise estimates of \( \phi \) in 3.3.2. The idea here is to see if noisy estimates of \( \phi \) contaminate the model. The results of this analysis are reported in table 3.2 and below is the associated graph of the marginal effect of competition:
The number of observations drops only marginally (773 drop from the sample). Comparing to the results in table 3.1, we find negligible qualitative differences. Point estimates and standard errors are essentially the same for all covariates except competition. The difference in the competition coefficients is reflected in the marginal impact. The marginal impact of competition in this specification is even stronger. The 95% confidence band does not include zero for values of \( c \) greater than approximately 0.8 (compared to approximately 0.95 with only positive estimates of \( \phi \)). Moreover, at the bottom 1% of \( c \) (approximately when \( c = 0.67 \)), the estimated marginal effect is noticeably above zero. When we consider only positive values of \( \phi \), the estimated marginal impact of competition at the bottom 1% (approximately when \( c = 0.77 \)) is close to zero.
3.7 Conclusion

The industrial organization literature on competition and innovation has developed tremendously since its inception. Simple theories on incumbency and the incentive to innovate have been replaced with more sophisticated models of inter-industry competition and R&D intensity. Likewise, simple cross-sectional studies on scale and R&D activity have been replaced with more sophisticated panel studies on market power and innovative output. Despite these advances, little attention has been given to the issue of measurement.

Standard measures of competition operate on the assumption that less market power, or the liberty with which a firm can price in excess of its marginal cost, can accurately proxy “more competitive.” I have argued that it cannot. Market power as it is defined in the literature is a profit margin. As such, it is not robust to cost-minimizing behavior, the economic incentive to cover fixed costs or the simple fact that superior performance will translate to higher returns. If anything, market power bears a closer resemblance to innovative prowess than to a comparative lack of competition.

What makes competition effective is its power to discipline. A textbook monopolist has little to no incentive to innovate: if it successfully innovates, it merely replaces itself; and if it fails, it faces a sunk cost. What would happen, though, if this monopolist had to suddenly compete? It would naturally consider the consequence of falling behind its competitor. This is the essence of why competition works—it forces companies to consider and work toward their survival. Such an observation is not new, and yet the literature continues to measure competition in ways that fail to capture this fundamental fact.

The primary objective of my research was to develop a simple framework where
the disciplinary mechanism of competition is explicit. To this end, I adopt a definition of “more competitive” advanced by Carl Shapiro: an industry is more competitive if a firm stands to lose greater profitable sales to its rivals should it offer inferior value to consumers. I apply this definition both theoretically and empirically.

My principal theoretical analysis operates under the assumption of stochastic innovation. Two firms choose their innovation strategies with the expectation that, should only one successfully innovate, the unsuccessful firm will have a portion of its profits captured by its rival. The greater the portion, the more competitive the market. This framework predicts a robust, positive relationship between competition and innovation.

I also address the empirical measurement of innovation. Recent research in empirical industrial organization has gravitated toward patents as a measure of innovation, replacing R&D. Patents are problematic for several reasons. First, the economic incentive to patent is not confined to protecting new, valuable ideas; companies also patent to fend off litigation or to launch a lawsuit themselves. Second, patents represent only a fraction of innovations; trade secret and lead-time to market are other ways to secure a competitive advantage. Third, specifically as it pertains to citation-weighted patent counts, the assumption of monotonicity between citations and value has been challenged by empirical evidence. R&D is another common measure of innovation; however, it has the disadvantage of being an input, not an output, to the innovative process. For these reasons, and the fact that innovation drives long-run economic growth, I measure innovation via productivity.

To test my model’s prediction of a positive, monotonic relationship between competition and innovation, I derive an empirical measure of competition that theoretically identifies how contestable a market is. I measure it as the elasticity of firm
market value with respect to sales, where sales proxy consumer value. It estimates how much market value a firm would have lost, in expectation, if it hypothetically generated less value for consumers.

An empirical test confirms a positive, monotonic relationship. I estimate a dynamic panel model with unobserved firm-level heterogeneity and a non-linear specification in competition. The sample used contains data on U.S. publicly traded manufacturing firms over the period 1962-2009. The results indicate that productivity not only increases with respect to competition, but in fact increases at the margin. In addition to this result, I also find evidence that recent market success, or at least the expectation of it, has a delayed negative impact on productivity, and that more competitors have a positive impact. Altogether, the results indicate that competition spurs innovation.

3.8 Future work

Several steps can be taken to further develop this framework. One, the theoretical model can be extended to a dynamic one. While the principle of “more competitive” is explicit in my model, it would be interesting to see how dynamics affect the competition-innovation relationship. Two, a different algorithm to estimate productivity, such as the one in Ackerberg et al. (2015), should be explored to assess the extent of bias—via collinearity in the first stage estimation—in the OP estimates. Three, a different set of data should be analyzed to test to the robustness of my results. My sample of data is large, but it is restricted to publicly traded manufacturing firms in the US, which is hardly a representative sample. And finally, modification of the productivity equation to account for product differentiation should be analyzed, as an underlying assumption of the OP method is that goods are homogeneous.
Table 3.1: Results with positive $\phi$

<table>
<thead>
<tr>
<th></th>
<th>HNR</th>
<th>FE</th>
<th>OLS</th>
<th>SGMM</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td>(1)</td>
<td>(2)</td>
</tr>
<tr>
<td>Productivity$_{t-1}$</td>
<td>0.483***</td>
<td>0.510***</td>
<td>0.382***</td>
<td>0.466***</td>
</tr>
<tr>
<td></td>
<td>(14.98)</td>
<td>(15.46)</td>
<td>(25.67)</td>
<td>(30.23)</td>
</tr>
<tr>
<td>Productivity$_{t-2}$</td>
<td>0.064***</td>
<td>0.075***</td>
<td>0.042***</td>
<td>0.097***</td>
</tr>
<tr>
<td></td>
<td>(4.07)</td>
<td>(4.56)</td>
<td>(3.53)</td>
<td>(7.62)</td>
</tr>
<tr>
<td>Productivity$_{t-3}$</td>
<td>0.034*</td>
<td>0.044**</td>
<td>0.043***</td>
<td>0.100***</td>
</tr>
<tr>
<td>Market value$_{t-1}$</td>
<td>-0.050</td>
<td>-0.050*</td>
<td>0.070***</td>
<td>0.047***</td>
</tr>
<tr>
<td></td>
<td>(−1.72)</td>
<td>(−2.23)</td>
<td>(12.44)</td>
<td>(8.45)</td>
</tr>
<tr>
<td>Market value$_{t-2}$</td>
<td>-0.069***</td>
<td>-0.070***</td>
<td>-0.063***</td>
<td>-0.073***</td>
</tr>
<tr>
<td></td>
<td>(−8.96)</td>
<td>(−9.04)</td>
<td>(−8.83)</td>
<td>(−10.01)</td>
</tr>
<tr>
<td>Market value$_{t-3}$</td>
<td>0.004</td>
<td>0.006</td>
<td>−0.019***</td>
<td>−0.001</td>
</tr>
<tr>
<td></td>
<td>(0.60)</td>
<td>(0.79)</td>
<td>(−3.41)</td>
<td>(−0.10)</td>
</tr>
<tr>
<td>Merge/acquire$_{t}$</td>
<td>−0.036</td>
<td>−0.033</td>
<td>−0.009</td>
<td>−0.024</td>
</tr>
<tr>
<td></td>
<td>(−1.61)</td>
<td>(−1.50)</td>
<td>(−0.46)</td>
<td>(−1.22)</td>
</tr>
<tr>
<td>Merge/acquire$_{t-1}$</td>
<td>0.001</td>
<td>0.006</td>
<td>0.028</td>
<td>0.022</td>
</tr>
<tr>
<td></td>
<td>(0.04)</td>
<td>(0.23)</td>
<td>(1.32)</td>
<td>(1.01)</td>
</tr>
<tr>
<td>Merge/acquire$_{t-2}$</td>
<td>0.051</td>
<td>0.058</td>
<td>0.047</td>
<td>0.043</td>
</tr>
<tr>
<td></td>
<td>(1.65)</td>
<td>(1.81)</td>
<td>(1.94)</td>
<td>(1.73)</td>
</tr>
<tr>
<td>Merge/acquire$_{t-3}$</td>
<td>−0.001</td>
<td>0.006</td>
<td>−0.004</td>
<td>−0.010</td>
</tr>
<tr>
<td></td>
<td>(−0.03)</td>
<td>(0.21)</td>
<td>(−0.18)</td>
<td>(−0.39)</td>
</tr>
<tr>
<td>Competition$_{t-1}$</td>
<td>0.200*</td>
<td>0.194*</td>
<td>0.051</td>
<td>−0.038</td>
</tr>
<tr>
<td></td>
<td>(2.53)</td>
<td>(2.47)</td>
<td>(1.23)</td>
<td>(−0.99)</td>
</tr>
<tr>
<td>Competition squared$_{t-1}$</td>
<td>0.391*</td>
<td>0.361*</td>
<td>0.277*</td>
<td>0.208*</td>
</tr>
<tr>
<td></td>
<td>(2.52)</td>
<td>(2.47)</td>
<td>(2.29)</td>
<td>(2.03)</td>
</tr>
<tr>
<td>Number of firms$_{t}$</td>
<td>0.115**</td>
<td>0.138**</td>
<td>0.027**</td>
<td>0.086***</td>
</tr>
<tr>
<td></td>
<td>(2.68)</td>
<td>(2.80)</td>
<td>(3.22)</td>
<td>(10.56)</td>
</tr>
<tr>
<td>Median industry return$_{t}$</td>
<td>0.098***</td>
<td>0.104***</td>
<td>0.182***</td>
<td>0.175***</td>
</tr>
<tr>
<td></td>
<td>(3.75)</td>
<td>(4.09)</td>
<td>(9.09)</td>
<td>(8.56)</td>
</tr>
</tbody>
</table>

Observations 33338 33338 36949 36949 36949 36949 36949 36949
Firms 2818 2818 3137 3137 3137 3137
Instruments 189 64
Hansen J 0.848 0.276
A-B 0.587 0.608
Table 3.2: Results with statistically significant $\phi$

<table>
<thead>
<tr>
<th></th>
<th>HNR</th>
<th>FE</th>
<th>OLS</th>
<th>SGMM</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td>(1)</td>
<td>(2)</td>
</tr>
<tr>
<td>Productivity$_{t-1}$</td>
<td>0.484***</td>
<td>0.512***</td>
<td>0.384***</td>
<td>0.468***</td>
</tr>
<tr>
<td></td>
<td>(14.64)</td>
<td>(15.20)</td>
<td>(25.39)</td>
<td>(29.99)</td>
</tr>
<tr>
<td>Productivity$_{t-2}$</td>
<td>0.060***</td>
<td>0.071***</td>
<td>0.039***</td>
<td>0.094***</td>
</tr>
<tr>
<td></td>
<td>(3.87)</td>
<td>(4.42)</td>
<td>(3.41)</td>
<td>(7.69)</td>
</tr>
<tr>
<td>Productivity$_{t-3}$</td>
<td>0.039**</td>
<td>0.050***</td>
<td>0.046***</td>
<td>0.104***</td>
</tr>
<tr>
<td></td>
<td>(2.79)</td>
<td>(3.50)</td>
<td>(4.45)</td>
<td>(9.59)</td>
</tr>
<tr>
<td>Market value$_{t-1}$</td>
<td>−0.053</td>
<td>−0.056*</td>
<td>0.070***</td>
<td>0.047***</td>
</tr>
<tr>
<td></td>
<td>(−1.79)</td>
<td>(−2.43)</td>
<td>(12.26)</td>
<td>(8.35)</td>
</tr>
<tr>
<td>Market value$_{t-2}$</td>
<td>−0.069***</td>
<td>−0.070***</td>
<td>−0.064***</td>
<td>−0.074***</td>
</tr>
<tr>
<td></td>
<td>(−8.83)</td>
<td>(−8.94)</td>
<td>(−8.75)</td>
<td>(−9.93)</td>
</tr>
<tr>
<td>Market value$_{t-3}$</td>
<td>0.003</td>
<td>0.005</td>
<td>−0.016**</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>(0.44)</td>
<td>(0.65)</td>
<td>(−3.20)</td>
<td>(0.05)</td>
</tr>
<tr>
<td>Merge/acquire$_t$</td>
<td>−0.035</td>
<td>−0.032</td>
<td>−0.009</td>
<td>−0.024</td>
</tr>
<tr>
<td></td>
<td>(−1.55)</td>
<td>(−1.45)</td>
<td>(−0.47)</td>
<td>(−1.20)</td>
</tr>
<tr>
<td>Merge/acquire$_{t-1}$</td>
<td>0.001</td>
<td>0.006</td>
<td>0.028</td>
<td>0.021</td>
</tr>
<tr>
<td></td>
<td>(0.03)</td>
<td>(0.21)</td>
<td>(1.28)</td>
<td>(0.95)</td>
</tr>
<tr>
<td>Merge/acquire$_{t-2}$</td>
<td>0.056</td>
<td>0.062</td>
<td>0.049*</td>
<td>0.045</td>
</tr>
<tr>
<td></td>
<td>(1.79)</td>
<td>(1.92)</td>
<td>(2.01)</td>
<td>(1.82)</td>
</tr>
<tr>
<td>Merge/acquire$_{t-3}$</td>
<td>−0.001</td>
<td>0.006</td>
<td>−0.005</td>
<td>−0.011</td>
</tr>
<tr>
<td></td>
<td>(−0.04)</td>
<td>(0.21)</td>
<td>(−0.21)</td>
<td>(−0.42)</td>
</tr>
<tr>
<td>Competition$_{t-1}$</td>
<td>0.319***</td>
<td>0.308***</td>
<td>0.054</td>
<td>−0.110*</td>
</tr>
<tr>
<td></td>
<td>(3.98)</td>
<td>(3.85)</td>
<td>(0.97)</td>
<td>(−2.08)</td>
</tr>
<tr>
<td>Competition squared$_{t-1}$</td>
<td>0.350*</td>
<td>0.320*</td>
<td>0.155</td>
<td>−0.089</td>
</tr>
<tr>
<td></td>
<td>(2.20)</td>
<td>(2.04)</td>
<td>(1.31)</td>
<td>(−0.77)</td>
</tr>
<tr>
<td>Number of firms$_{t}$</td>
<td>0.113***</td>
<td>0.135**</td>
<td>0.027**</td>
<td>0.086***</td>
</tr>
<tr>
<td></td>
<td>(2.62)</td>
<td>(2.73)</td>
<td>(3.29)</td>
<td>(10.61)</td>
</tr>
<tr>
<td>Median industry return$_t$</td>
<td>0.084**</td>
<td>0.088***</td>
<td>0.176***</td>
<td>0.167***</td>
</tr>
<tr>
<td></td>
<td>(3.12)</td>
<td>(3.37)</td>
<td>(8.68)</td>
<td>(8.17)</td>
</tr>
<tr>
<td>Observations</td>
<td>32595</td>
<td>32595</td>
<td>36217</td>
<td>36217</td>
</tr>
<tr>
<td>Firms</td>
<td>2766</td>
<td>2766</td>
<td>3086</td>
<td>3086</td>
</tr>
<tr>
<td>Instruments</td>
<td>186</td>
<td>64</td>
<td>190</td>
<td>68</td>
</tr>
<tr>
<td>Hansen J</td>
<td>0.770</td>
<td>0.281</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>A-B</td>
<td>0.704</td>
<td>0.737</td>
<td>0.241</td>
<td>0.282</td>
</tr>
</tbody>
</table>
Table 3.3: Results using different instrument lag lengths

<table>
<thead>
<tr>
<th></th>
<th>Lag(5)</th>
<th>Lag(4)</th>
<th>Lag(3)</th>
<th>Lag(2)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Productivity</strong>&lt;sub&gt;t-1&lt;/sub&gt;</td>
<td>0.468***</td>
<td>0.472***</td>
<td>0.475***</td>
<td>0.483***</td>
</tr>
<tr>
<td></td>
<td>(14.43)</td>
<td>(14.68)</td>
<td>(14.87)</td>
<td>(14.98)</td>
</tr>
<tr>
<td><strong>Productivity</strong>&lt;sub&gt;t-2&lt;/sub&gt;</td>
<td>0.062****</td>
<td>0.065***</td>
<td>0.063***</td>
<td>0.064***</td>
</tr>
<tr>
<td></td>
<td>(3.88)</td>
<td>(4.11)</td>
<td>(3.99)</td>
<td>(4.07)</td>
</tr>
<tr>
<td><strong>Productivity</strong>&lt;sub&gt;t-3&lt;/sub&gt;</td>
<td>0.035*</td>
<td>0.033*</td>
<td>0.035*</td>
<td>0.034*</td>
</tr>
<tr>
<td></td>
<td>(2.51)</td>
<td>(2.36)</td>
<td>(2.54)</td>
<td>(2.41)</td>
</tr>
<tr>
<td><strong>Market value</strong>&lt;sub&gt;t-1&lt;/sub&gt;</td>
<td>−0.036</td>
<td>−0.041</td>
<td>−0.043</td>
<td>−0.050</td>
</tr>
<tr>
<td></td>
<td>(−1.53)</td>
<td>(−1.72)</td>
<td>(−1.67)</td>
<td>(−1.72)</td>
</tr>
<tr>
<td><strong>Market value</strong>&lt;sub&gt;t-2&lt;/sub&gt;</td>
<td>−0.070***</td>
<td>−0.070***</td>
<td>−0.069***</td>
<td>−0.069***</td>
</tr>
<tr>
<td></td>
<td>(−9.07)</td>
<td>(−9.05)</td>
<td>(−8.95)</td>
<td>(−8.96)</td>
</tr>
<tr>
<td><strong>Market value</strong>&lt;sub&gt;t-3&lt;/sub&gt;</td>
<td>0.005</td>
<td>0.005</td>
<td>0.004</td>
<td>0.004</td>
</tr>
<tr>
<td></td>
<td>(0.69)</td>
<td>(0.64)</td>
<td>(0.57)</td>
<td>(0.60)</td>
</tr>
<tr>
<td><strong>Merge/acquire</strong>&lt;sub&gt;t&lt;/sub&gt;</td>
<td>−0.029</td>
<td>−0.032</td>
<td>−0.033</td>
<td>−0.036</td>
</tr>
<tr>
<td></td>
<td>(−1.31)</td>
<td>(−1.43)</td>
<td>(−1.48)</td>
<td>(−1.61)</td>
</tr>
<tr>
<td><strong>Merge/acquire</strong>&lt;sub&gt;t-1&lt;/sub&gt;</td>
<td>−0.003</td>
<td>−0.002</td>
<td>0.001</td>
<td>0.001</td>
</tr>
<tr>
<td></td>
<td>(−0.12)</td>
<td>(−0.09)</td>
<td>(0.05)</td>
<td>(0.04)</td>
</tr>
<tr>
<td><strong>Merge/acquire</strong>&lt;sub&gt;t-2&lt;/sub&gt;</td>
<td>0.052</td>
<td>0.050</td>
<td>0.055</td>
<td>0.051</td>
</tr>
<tr>
<td></td>
<td>(1.70)</td>
<td>(1.62)</td>
<td>(1.77)</td>
<td>(1.65)</td>
</tr>
<tr>
<td><strong>Merge/acquire</strong>&lt;sub&gt;t-3&lt;/sub&gt;</td>
<td>0.003</td>
<td>0.003</td>
<td>0.003</td>
<td>−0.001</td>
</tr>
<tr>
<td></td>
<td>(0.12)</td>
<td>(0.12)</td>
<td>(0.12)</td>
<td>(−0.03)</td>
</tr>
<tr>
<td><strong>Competition</strong>&lt;sub&gt;t-1&lt;/sub&gt;</td>
<td>0.185*</td>
<td>0.182*</td>
<td>0.198*</td>
<td>0.200*</td>
</tr>
<tr>
<td></td>
<td>(2.37)</td>
<td>(2.33)</td>
<td>(2.51)</td>
<td>(2.53)</td>
</tr>
<tr>
<td><strong>Competition squared</strong>&lt;sub&gt;t-1&lt;/sub&gt;</td>
<td>0.367*</td>
<td>0.380*</td>
<td>0.392*</td>
<td>0.391*</td>
</tr>
<tr>
<td></td>
<td>(2.32)</td>
<td>(2.41)</td>
<td>(2.50)</td>
<td>(2.52)</td>
</tr>
<tr>
<td><strong>Number of firms</strong>&lt;sub&gt;t&lt;/sub&gt;</td>
<td>0.094*</td>
<td>0.095*</td>
<td>0.098*</td>
<td>0.115**</td>
</tr>
<tr>
<td></td>
<td>(2.44)</td>
<td>(2.43)</td>
<td>(2.45)</td>
<td>(2.68)</td>
</tr>
<tr>
<td><strong>Median industry return</strong>&lt;sub&gt;t&lt;/sub&gt;</td>
<td>0.101***</td>
<td>0.099***</td>
<td>0.096***</td>
<td>0.098***</td>
</tr>
<tr>
<td></td>
<td>(4.01)</td>
<td>(3.91)</td>
<td>(3.74)</td>
<td>(3.75)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>33338</th>
<th>33338</th>
<th>33338</th>
<th>33338</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Observations</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Firms</strong></td>
<td>2818</td>
<td>2818</td>
<td>2818</td>
<td>2818</td>
</tr>
<tr>
<td><strong>Instruments</strong></td>
<td>198</td>
<td>195</td>
<td>192</td>
<td>189</td>
</tr>
<tr>
<td><strong>Hansen J</strong></td>
<td>0.212</td>
<td>0.599</td>
<td>0.707</td>
<td>0.848</td>
</tr>
<tr>
<td><strong>A-B</strong></td>
<td>0.666</td>
<td>0.455</td>
<td>0.627</td>
<td>0.587</td>
</tr>
</tbody>
</table>
Appendix A

Miscellaneous tables and figures
Figure A.1: Industry-level Plots: Time series plots by industry (1962–2009)

Figure A.2: Summary of Industry Productivity and Competition Calculations

<table>
<thead>
<tr>
<th>Industry (Manufacturing)</th>
<th>NAICS</th>
<th>Labor input elasticity</th>
<th>Std. Err.</th>
<th>Capital input elasticity</th>
<th>Std. Err.</th>
<th>Returns to scale</th>
<th>Competition</th>
<th>Number of firms</th>
<th>HHI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Food</td>
<td>311</td>
<td>0.67</td>
<td>0.07</td>
<td>0.37</td>
<td>0.08</td>
<td>1.03</td>
<td>2.21</td>
<td>26</td>
<td>1279</td>
</tr>
<tr>
<td>Beverages and Tobacco</td>
<td>312</td>
<td>0.60</td>
<td>0.08</td>
<td>0.23</td>
<td>0.10</td>
<td>0.83</td>
<td>2.53</td>
<td>15</td>
<td>1513</td>
</tr>
<tr>
<td>Textile Mills</td>
<td>313</td>
<td>0.70</td>
<td>0.10</td>
<td>0.19</td>
<td>0.06</td>
<td>0.89</td>
<td>2.24</td>
<td>24</td>
<td>1609</td>
</tr>
<tr>
<td>Textile Products Mills</td>
<td>314</td>
<td>0.82</td>
<td>0.09</td>
<td>0.08</td>
<td>0.09</td>
<td>0.90</td>
<td>7.45</td>
<td>6</td>
<td>5148</td>
</tr>
<tr>
<td>Wood Products</td>
<td>321</td>
<td>0.87</td>
<td>0.07</td>
<td>0.33</td>
<td>0.10</td>
<td>1.20</td>
<td>2.35</td>
<td>33</td>
<td>1644</td>
</tr>
<tr>
<td>Paper</td>
<td>322</td>
<td>0.70</td>
<td>0.06</td>
<td>0.37</td>
<td>0.09</td>
<td>1.07</td>
<td>2.96</td>
<td>33</td>
<td>858</td>
</tr>
<tr>
<td>Printing</td>
<td>323</td>
<td>0.68</td>
<td>0.11</td>
<td>0.31</td>
<td>0.13</td>
<td>0.99</td>
<td>1.10</td>
<td>10</td>
<td>3241</td>
</tr>
<tr>
<td>Petroleum and Coal</td>
<td>324</td>
<td>0.44</td>
<td>0.04</td>
<td>0.29</td>
<td>0.08</td>
<td>0.72</td>
<td>1.47</td>
<td>32</td>
<td>1030</td>
</tr>
<tr>
<td>Chemical</td>
<td>325</td>
<td>0.63</td>
<td>0.04</td>
<td>0.32</td>
<td>0.05</td>
<td>0.95</td>
<td>1.40</td>
<td>171</td>
<td>351</td>
</tr>
<tr>
<td>Plastics and Rubber</td>
<td>326</td>
<td>0.66</td>
<td>0.05</td>
<td>0.36</td>
<td>0.07</td>
<td>1.02</td>
<td>1.34</td>
<td>39</td>
<td>2183</td>
</tr>
<tr>
<td>Nonmetallic Mineral</td>
<td>327</td>
<td>0.59</td>
<td>0.10</td>
<td>0.22</td>
<td>0.10</td>
<td>0.81</td>
<td>3.02</td>
<td>12</td>
<td>1947</td>
</tr>
<tr>
<td>Primary Metal</td>
<td>331</td>
<td>0.72</td>
<td>0.07</td>
<td>0.06</td>
<td>0.10</td>
<td>0.78</td>
<td>1.24</td>
<td>24</td>
<td>1890</td>
</tr>
<tr>
<td>Fabricated Metal</td>
<td>332</td>
<td>0.71</td>
<td>0.04</td>
<td>0.32</td>
<td>0.05</td>
<td>1.03</td>
<td>3.01</td>
<td>52</td>
<td>752</td>
</tr>
<tr>
<td>Machinery</td>
<td>333</td>
<td>0.83</td>
<td>0.03</td>
<td>0.21</td>
<td>0.02</td>
<td>1.04</td>
<td>2.29</td>
<td>131</td>
<td>729</td>
</tr>
<tr>
<td>Computer and Electronic</td>
<td>334</td>
<td>0.73</td>
<td>0.03</td>
<td>0.34</td>
<td>0.04</td>
<td>1.07</td>
<td>3.22</td>
<td>421</td>
<td>359</td>
</tr>
<tr>
<td>Electrical Equipment</td>
<td>335</td>
<td>0.84</td>
<td>0.05</td>
<td>0.21</td>
<td>0.06</td>
<td>1.06</td>
<td>1.91</td>
<td>53</td>
<td>1987</td>
</tr>
<tr>
<td>Transportation Equipment</td>
<td>336</td>
<td>0.80</td>
<td>0.04</td>
<td>0.19</td>
<td>0.04</td>
<td>0.99</td>
<td>1.18</td>
<td>65</td>
<td>941</td>
</tr>
<tr>
<td>Furniture and Related</td>
<td>337</td>
<td>0.86</td>
<td>0.06</td>
<td>0.15</td>
<td>0.11</td>
<td>1.01</td>
<td>4.87</td>
<td>15</td>
<td>1884</td>
</tr>
<tr>
<td>Miscellaneous</td>
<td>339</td>
<td>0.62</td>
<td>0.05</td>
<td>0.43</td>
<td>0.05</td>
<td>1.06</td>
<td>3.30</td>
<td>83</td>
<td>842</td>
</tr>
</tbody>
</table>
Figure A.3 : Industry-level Plots: Innovation against competition (standardized)
Figure A.4: Industry-level Plots: Competition against standard measures of market power and market structure (standardized)
Bibliography


