1 A Divorce between Theory and Empirics

Looking back to the mid-nineties, a curious economic observer seeking to form a coherent view on what impact competition policy had on growth would find herself discouraged by the lack of consensus on the subject. Existing theories were in sharp juxtaposition to both the common wisdom and empirical evidence. On the one hand, she would find that the leading theoretical models in industrial organization or in growth theory predicted that more intense product market competition discourages innovation and growth as it reduces the rents from innovating (the argument used by the Bill Gateses of this world to oppose antitrust action). On the other hand, the common view, dating back to Adam Smith and put forward more recently by economists including Michael Porter, was that competition enhances growth because it exerts pressure on firms to cut costs, reduce slack, and innovate in order to maintain market position, by introducing new products or new production processes. These beliefs resulted in wide-ranging policy initiatives aimed at facilitating competition by promoting openness, free trade, free entry by foreign investors, and monetary integration across the world.

Now why should a serious observer believe the common wisdom if it is not supported by empirical evidence? If she were truly serious, our observer would look to recent econometric studies on the subject, hoping that they would confirm the theories and prove the common wisdom wrong. However, the "state-of-the-art" microeconometric studies of the 1990s would add to her confusion, by pointing to an unambiguously positive correlation between productivity growth and various measures of the intensity of product market competition. In the face of this evidence, she could only be left feeling discomfort with the theory. Yet would she be right to throw out the theory? Or, like the proverbial baby in the bathwater, should more care be taken?

In this chapter, we provide an account of the divide between applied theorists and empiricists in their approach to competition and its effects on growth. In doing so, we try to uncover missing or embryonic elements on either side that could hint at the possibility of a subsequent reconciliation between the different views.

1.1 The Dominant Theories by the Early 1990s

To gain an understanding of the dominant view on competition, innovation, and entry, our observer would naturally consult Tirole's (1988) reference textbook on the theory of industrial organization, then look at the more recent endogenous growth literature (e.g., Romer 1990; Aghion and Howitt 1992; Grossman and Helpman 1991). In this section, we provide a brief account of what our observer would have learned from her theoretical exploration.

1.1.1 The IO Models of Product Differentiation and Price Competition, and the Schumpeterian Effect of Competition

The two leading models of price competition and product differentiation in theoretical IO, are the Hotelling linear model (and the circular version of that model by Salop (1977)) and the symmetric model of monopolistic competition by Dixit and Stiglitz (1977). This latter model has been the template for Romer's (1990) model of endogenous growth with increasing product variety. Both models are described in detail in chapter 7 of Tirole 1988, and they deliver the same prediction: More intense product market competition reduces the rents of those firms that successfully enter the market, and therefore it discourages firms from entering in the first place. Entry in these models is what captures the notion of innovation.

The Circular Model Salop's circular model, shown in figure 1.1, is one where the market is represented by a circle of unit length, on which firms locate evenly. Thus, if there are n firms in the market, the "distance" between two neighboring firms is $1/n$. Consumers are uniformly distributed over the circle, and they must incur a transportation cost t per unit of length they travel through. This parameter t captures the extent of product market competition. The higher the t, the more costly it is for a consumer to shift from one firm to another, and therefore the less an individual firm on the circle has to worry about the risk that consumers located in her immediate neighborhood be competed

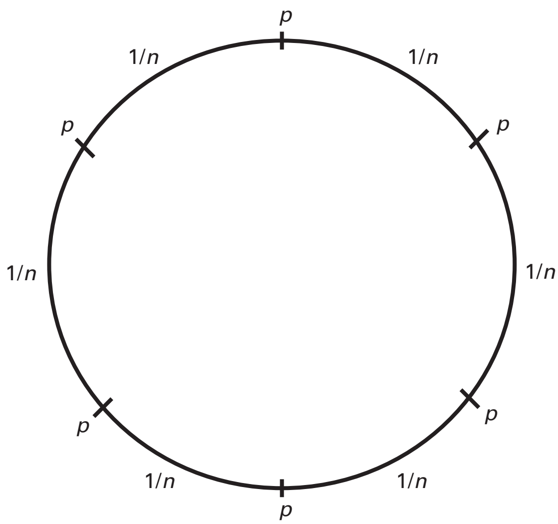

Figure 1.1
Salop circle model

away by other firms, even if the firm sets a price close to its unconstrained monopoly price. Thus, a higher t corresponds to a lower degree of product market competition.

The timing of the model is as follows. In a first stage, firms with identical unit production cost c decide whether or not to *enter* the market, where entry involves a fixed cost f. Think of entry in this model as capturing the innovation decision of a firm. More entry corresponds to more (product) innovation. In a second stage, those firms that have entered the market compete in price, that is, engage in Bertrand competition. Our main question is how an increase in product market competition, modeled as a reduction in t, affects innovation, measured by entry or the equilibrium number of firms in the market. The answer turns out to be straightforward and unambiguous, namely, that increased product market competition discourages entry.

To understand why more formally, we solve the model by backward induction, first solving for the Nash equilibrium of the price competition game for a given number of firms n in the market, then moving back to the entry stage and solving for the equilibrium value of n that makes firms just indifferent between incurring the entry cost f and staying out of the market. Here we implicitly assume free entry, which means that firms keep entering the circle until the marginal firm finds it unprofitable to pay the entry cost f, given the number of firms already on the market.

Price competition: We restrict attention to symmetric Nash equilibria where all firms charge the same price p in equilibrium. If firm i chooses price p_i and all other firms have chosen price p, then the consumers who will be indifferent between purchasing from firm i or its neighbor are located at distance x from firm i on either side, such that

$$p_i + tx = p + t(1/n - x).$$

This implies that the total demand for firm i's product will be

$$D(p_i, p) = 2x = \frac{t/n + p - p_i}{t}.$$

Firm i will thus react to price p by its competitors, by choosing p_i so as to maximize its current profit

$$\pi(p_i, p) = (p_i - c)D(p_i, p).$$

Solving for p_i by taking the first-order condition for this maximization, then using the fact that in a symmetric Nash equilibrium of this price competition game,

$$p_i = p,$$

we obtain the following equilibrium price and profit flows:

$$p^* = \frac{t}{n} + c$$

and

$$\pi^*(n) = \frac{t}{n^2}.$$

Not surprisingly, an increase in product market competition reduces the equilibrium level of profits for firms in the market. In the absence of product differentiation, that is, when $t = 0$, we are back to the traditional Bertrand competition case where profits are competed down to zero.

Entry: Moving back to the initial entry stage, the equilibrium number of firms n^* is determined by the free-entry condition

$$\pi^*(n) = f,$$

which immediately yields

$$n^* = \sqrt{t/f}.$$

In particular, an increase in product market competition, modeled as a reduction in transportation costs, discourages entry by reducing post-entry rents. As Dasgupta and Stiglitz (1980) have suggested, ex post competition drives out ex ante competition. We refer to this as the Schumpeterian effect of product market competition.

The Dixit-Stiglitz Model A similar conclusion obtains in the Dixit-Stiglitz model of product differentiation, where consumers all share the same utility for the differentiated goods, of the form

$$u(q_1, \ldots, q_n) = \left(\sum_{j=1}^{n} q_j^{\alpha} \right)^{1/\alpha} \tag{1.1}$$

where q_i is quantity of good i consumed. In this model, product market competition is captured by the parameter α, with a higher α corresponding to a higher degree of substitutability between the differentiated products, and therefore to a higher degree of competition between the firms that produce them. We again assume symmetry among differentiated goods producers, with all of them facing the same unit production cost c. Then, we can analyze the same two-stage game as before, where differentiated goods producers first decide whether to pay a fixed entry fee f and enter the market, and then those who entered the market compete in price.

Price competition: We take the total number of firms to be large, so that an individual firm i takes the total amount of consumption

$$\sum_{j=1}^{n} q_j^{\alpha}$$

as given when choosing its price p_i. The inverse demand function for product i is obtained by equating the price p_i of good i to the marginal utility of that good, namely,

$$p_i = \frac{\partial u}{\partial q_i} = q_i^{\alpha-1} \left(\sum_{i=1}^{n} q_i^{\alpha} \right)^{(1/\alpha)-1}, \tag{1.2}$$

so that when we solve for the quantity demanded we get

$$D(p_i) = q_i = k p_i^{-1/(1-\alpha)},$$

where

$$k = \left(\sum_{j=1}^{n} q_j^{\alpha} \Big/ w \right)^{1/(\alpha-1)}$$

is treated as a constant by each individual firm i for n sufficiently large, and where w denotes the wealth of the representative consumer.

Thus firm i chooses the price p_i that maximizes $(p_i - c_i)p_i^{1/(\alpha-1)}$, which implies

$$p_i = \frac{c}{\alpha},$$

that is, price equals marginal cost scaled up by the degree of substitutability of products. The higher the degree of substitutability, the lower the price. Now, substituting for p_i into equation (2.2), and using the fact that all firms produce the same quantity q in a symmetric Nash equilibrium, we obtain the equilibrium profit

$$\pi^*(n) = (1-\alpha)\frac{w}{n}.$$

Entry: Once again, the equilibrium level of entry n^* is simply determined by the free-entry condition

$$\pi^*(n) = f.$$

This yields the simple expression

$$n^* = \frac{(1-\alpha)w}{f},$$

which again shows that an increase in product market competition, here modeled as an increase in the substitutability between the differentiated goods as measured by α, reduces post-entry rents and therefore discourages entry (or innovation). Thus, we again obtain an unambiguously negative Schumpeterian effect of product market competition on innovation.

1.1.2 Two Attempts at Generating a Positive Effect of Competition on Entry or Innovation

Our observer would thus walk away from her exploration of this early IO theory with the idea that more intense competition discourages entry, because it reduces post-entry rents. She would be wrong, however, to think that all IO models of competition and entry or competition and innovation predict a negative impact of competition. In particular, she will have missed two important insights from IO models, namely, the interplay between rent dissipation and preemption incentives, and the differences between vertical (i.e., quality improving) and horizontal innovations. Those insights, which we briefly spell out in this section, will prove to be useful when, in subsequent chapters, we try to reconcile theory with empirical evidence on the relationship between competition and growth.

The Rent Dissipation Effect Chapters 8 and 10 of Tirole 1988 present closely related models of preemption and innovation,[3] which suggest a positive effect of product market competition on innovation. Suppose, for example, that an incumbent firm is engaged in a race with a potential entrant for a new innovation that will reduce costs. Who will invest more research and development (R&D) resources in the race, the incumbent or the potential entrant? The answer turns out to be ambiguous, and it relies on the trade-off between two opposite effects: a *rent dissipation* effect and a *replacement* effect. The replacement effect, uncovered by Kenneth Arrow in 1962, refers to the fact that, by innovating, the incumbent monopolist replaces her own rents, whereas the potential entrant has no preexisting rents to replace. Everything else remaining equal, this effect will induce the entrant to invest more in the race than the incumbent firm will. On the other hand, the rent dissipation effect refers to the fact that the incumbent may lose more by letting the entrant win the race (she dissipates the difference between her current monopoly rents and the duopoly rents if the entrant innovates) than the potential entrant does by letting the incumbent win the race (he loses the difference between what may be at best duopoly rents if he had won the race and zero if the incumbent wins). The rent dissipation effect may or may not counteract the replacement effect. If it does, then the incumbent ends up investing more in the race than does the potential entrant.

How does this relate to product market competition? To get some preliminary intuition, let us go back to our first model of product differentiation, and consider a firm located at one extreme of a linear city, which faces the risk that a second firm will enter and locate at the other extreme of the city. The linear city is like the circular city analyzed earlier, except that consumers are now uniformly distributed on a segment of length one. We still denote by t the unit transport cost.

The only way the incumbent firm can prevent entry is to use its incumbency advantage and to build a second plant at the other extreme of the line before the entrants does. Suppose that the second firm (the potential entrant) observes what the incumbent does before deciding whether or not to enter, and also whether entry involves a positive sunk cost, f. Then, anticipating (undifferentiated) Bertrand competition with the second plant, and therefore zero profits in case it enters, the second firm will not enter, as entry will lead to a net loss of f.

Of course, the second firm could enter into a race with the incumbent in order to arrive first at the other location. However, there is asymmetry between the two firms' incentives to invest in that race. On the one hand, the incumbent firm will lose

$$\pi^m - \pi^d$$

per unit of time if entry occurs, where π^m denotes the incumbent's monopoly profit flow if she wins the race,[4] and π^d denotes the equilibrium duopoly profit of each firm if the entrant succeeds in locating first at the other extreme of the line. Thus, the incumbent firm's incentive to invest in product innovation at the other end of the segment is proportional to $\pi^m - \pi^d$.

On the other hand, by investing in innovation, the potential entrant raises the chance of moving from zero to π^d, and thus its incentive to invest in innovation is proportional to π^d. Note that the potential entrant, like firms in the previous models of monopolistic competition and entry, moves from zero to something positive, but the positive amount is decreasing with competition. However, unlike in the previous models, the incumbent firm starts with positive profits when deciding whether or not to innovate. This, in turn, makes a big difference, as we will see repeatedly in this book.

The IO literature emphasizes the comparison between $\pi^m - \pi^d$ and π^d, and the fact that when competition generates enough rent dissipation (reduces π^d sufficiently), then

$$\pi^m - \pi^d > \pi^d,$$

so that the incumbent is more likely to win the race and thereby persist as a monopoly.

However, it does not consider the effect of an increase in product market competition (i.e., of a reduction in the transport cost t) on π^m and π^d. Clearly, the entrant responds negatively to an increase in product market competition, as his post-entry profit π^d decreases when t decreases (this is the Schumpeterian effect emphasized earlier). On the other hand, the incumbent may respond positively to higher competition, insofar as π^d decreases more with competition than π^m does, so that the rent dissipation $\pi^m - \pi^d$ goes up when t goes down.

Much of the analysis in this book revolves around these two effects, first within a particular sector, and then across sectors with different technological characteristics, and their impact on the magnitude of cost or quality differences between incumbent and entrant. This brings us to a second important extension of the above models of price competition and entry.

The Importance of Vertical Differentiation Let us go back to the circular model, but now suppose that some firms have higher unit costs than others. Thus firms are not only horizontally differentiated along the circle, but also vertically differentiated by their costs. In this case, as shown in Aghion and Schankerman 2003, more intense product market competition, modeled again as a reduction in the unit transport cost t, can enhance "innovations" through several channels that counteract the negative effect pointed out previously. First, by increasing the market share of low-cost firms at the expense of high-cost firms (this is referred to as the "election effect" of product market competition), more intense competition may end up encouraging entry by low-cost firms (especially if potential low-cost entrants are far less numerous than high-cost entrants). Second, and again because it increases the market share of low-cost firms relative to high-cost firms, more intense competition will induce high-cost firms to invest in "restructuring" in order to become low-cost firms themselves. Note that such an investment amounts to a quality-improving innovation that allows the high-cost firm to suffer less from more intense competition. This type of effect will also play an important role in subsequent chapters of this book.

1.1.3 The Endogenous Growth Paradigm

Main Idea Reading around the literature more broadly, our explorer would find that the prediction that product market competition has an unambiguously negative effect on entry or innovation is shared by the models of endogenous technical change in growth theory (e.g., Romer 1990; Aghion and Howitt 1992; Grossman and Helpman 1991). In all of these models, an increase in product market competition, or in the rate of imitation, has a negative effect on productivity growth by reducing the monopoly rents that reward new innovation. This discourages firms from engaging in R&D activities, thereby lowering the innovation rate and therefore also the rate of long-run growth, which in these models is proportional to the innovation rate. In the product variety framework of Romer (1990), this property is directly inherited from the Dixit-Stiglitz model upon which this model is built. But the same effect is also at work in the Schumpeterian (or quality-ladder) models of Aghion and Howitt (1992) and Grossman and Helpman (1991). These two models predict that property right protection is growth-enhancing, however, for exactly the same reason they also predict that competition policy is unambiguously detrimental to growth: Patent protection protects monopoly rents from innovation, whereas increased product market competition destroys these rents. Thus, if we were to take these models at face value when making policy prescriptions, we would never advocate that patent policy and antitrust be pursued at the same time, at least not from the point of view of promoting dynamic efficiency.[5]

Here we present a simplified version of the Schumpeterian growth model with quality-improving innovations. This serves as a basis for the theoretical extensions we will present in later chapters of this book and provide a framework in which the tension between theory and evidence can be reconciled.

A Benchmark Model of Innovation and Productivity Growth Consider an economy with a final good, y, and a continuum of intermediate inputs indexed by $i \in [0,1]$. Time is discrete, indexed by $t = 1, 2, \ldots, T$. One final good is produced competitively using a continuum of mass 1 of intermediate inputs according to the constant returns to scale production function:

$$y_t = \int_0^1 A_t(i)^{1-\alpha} x_t(i)^{\alpha}\, di, \tag{1.3}$$

where each $x_t(i)$ is the flow of intermediate input i used at date t, and $A_t(i)$ is a productivity variable that measures the quality of the input. This variable will grow over time as a result of quality-improving innovations. The final good is used in turn for consumption, research, and production of the intermediate inputs. For notational simplicity, we omit the arguement i except when it is necessary.

Each intermediate sector is monopolized by an incumbent producer (the incumbent innovator in that sector) who can produce the leading-edge version of input i at a constant marginal cost of one unit of the final good. Each individual producer lives for one period only and therefore maximizes short-run profits. But she faces a competitive fringe of imitators who can produce the same input at a constant marginal cost $\chi > 1$. The parameter χ is an inverse measure of the degree of product market competition or imitation in the economy: The higher is χ, the greater the innovators' market power and thus the lower the degree of competition.[6] The incumbent producer is forced to charge a limit price (in terms of the final good, our numeraire) equal to

$$p_t = \chi.$$

to prevent the fringe from stealing her market.

Because the final-good-producing sector is competitive, price is also equal to marginal productivity:

$$p_t = \partial y_t / \partial x_t(i) = \alpha (x_t(i)/A_t(i))^{\alpha - 1}.$$

Equating the two expressions for the price, we get

$$x_t(i) = \left(\frac{\chi}{\alpha}\right)^{1/(\alpha-1)} A_t(i),$$

so that the equilibrium monopoly rent of the incumbent producer in sector i is equal to

$$\pi_t(i) = (p_t - 1)x_t(i) = \delta(\chi)A_t(i),$$

where

$$\delta(\chi) \equiv (\chi - 1)(\chi/\alpha)^{1/(\alpha-1)}.$$

Innovations in sector i at the beginning of period t result in an improved version of the corresponding intermediate input. Namely, an innovation at t multiplies the preexisting productivity parameter $A_{t-1}(i)$ by a factor $\gamma > 1$.

Innovations in turn result from research z. By incurring an effort cost

$$c_{ti}(z) = \frac{1}{2} A_{t-1}(i) z^2$$

at the beginning of the period, some individual in sector i can become the new "leading-edge" producer of the intermediate input with probability λz. The payoff to research in sector i is the prospect of the monopoly rent $\pi_t(i)$ if the research succeeds in producing an innovation.

Assuming that all individuals can imitate the current technology after one period, this implies that a non-innovating incumbent makes no current profit, and that the monopoly rent of an innovating producer lasts for one period only.

Assume the time period is short enough that we may ignore the possibility of more than one successful innovator in the same sector. Then

$$A_t(i) = \begin{cases} \gamma A_{t-1}(i) & \text{with probability } \lambda n \\ A_{t-1}(i) & \text{with probability } 1 - \lambda n \end{cases} \tag{1.4}$$

where n is the equilibrium R&D investment in any sector i. The average growth rate is then simply given by

$$g = E(\ln A_t(i) - \ln A_{t-1}(i)) = \lambda n \ln \gamma$$

in an equilibrium where productivity-adjusted research is a constant n.

Now, the optimal R&D investment is the one that maximizes expected profits minus costs, namely,

$$\max_z \left\{ \lambda z \pi_t(i) - \frac{1}{2} A_{t-1}(i) z^2 \right\},$$

with first-order condition

$$\begin{aligned} z = n &= \lambda(\pi_t(i)/A_{t-1}(i)) \\ &= \lambda\gamma(\pi_t(i)/A_t(i)), \end{aligned}$$

or equivalently,

$$n = \lambda\gamma(\chi - 1)(\chi/\alpha)^{1/(\alpha-1)}.$$

The corresponding average rate of productivity growth is simply

$$g = \lambda^2\gamma(\chi - 1)(\chi/\alpha)^{1/(\alpha-1)} \ln \gamma.$$

In particular, productivity growth is decreasing with the degree of product market competition (or with the degree of imitation), as inversely measured by χ. Thus, as we stressed at the beginning of this section, patent protection (or, more generally, better protection of intellectual property rights) will enhance growth by increasing χ and therefore increasing potential rewards from innovation. However, pro-competition policies will tend to discourage innovation and growth by reducing χ, and thereby forcing incumbent innovators to charge a lower limit price.

Our observer may ask: How come we have a model with vertical innovation that nevertheless delivers the same prediction as the Salop and Dixit-Stiglitz models of horizontal differentiation and entry? At the same time, our reader will notice that, as in these other two models, innovation is always performed by outsiders, that is, by firms that make no profit before they innovate. This is in contrast with the preemption or restructuring models mentioned in the previous section. Also, note that all the researchers racing for the new innovation have access to the same R&D technology and they achieve the same productivity level if they successfully innovate, unlike high- versus low-cost entrants in the Aghion-Schankerman model. These considerations will suggest natural ways of extending the model so as to reconcile theory and empirics, as we will see in subsequent chapters.

1.2 The Evidence Contradicts the Theory

Before considering these theoretical extensions, our explorer may well decide to look to the empirical literature. Does it support the theory as laid out so far? Does it offer alternative avenues for investigation? The empirical literature linking product market competition to innovation and productivity growth predates the theoretical literature by several decades. Thus, our observer starts by looking back to the empirical literature of the mid-sixties, pioneered by Scherer, then jumps ahead (skipping volumes of work) twenty-five years to the more recent microecometric literature of the 1990s. And, as before, our explorer may spot a number of limitations to the various pieces of work, which upon closer examination might suggest ways to reconcile the theory and the evidence.

1.2.1 The Early Literature

A large early empirical literature, inspired by Schumpeter (1943), considered the cross-sectional relationship between innovation and firm size or market concentration.[7] Many studies found that larger firms (either measured by size or market share) were also more innovative (or spent more on R&D). Figure 1.2 shows this pattern, which is seen across a large number of datasets. Here we have graphed the average number of patents taken out at the U.S. Patent Office by firms listed on the London Stock Exchange. On the x-axis we have ranked firms by their size decile; the smallest firms are located in the first decile and

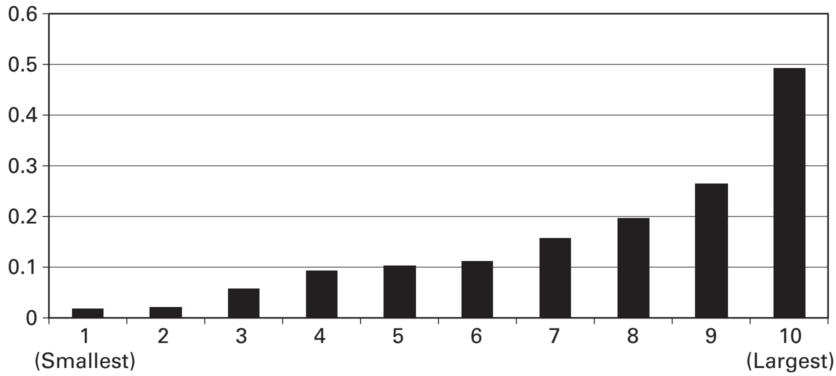

Figure 1.2
Patents by firm size decile

the largest firms in the tenth decile. The graph clearly shows that the bulk of patenting is done by larger firms.[8]

Scherer's early empirical work[9] showed that there was a relationship between firm patenting activity and firm size in the cross section. For example, Scherer (1965a) used patents data on Fortune 500 firms in 1959 and regressed this on sales in 1955. He found a positive relationship. However, interestingly, he also found that when he allowed for non-linearities these suggested a diminishing impact at larger sizes. We will return to these non-linearities in later chapters. Scherer (1965a) also investigated the relationship between four firm concentration indices and patenting activity and finds no significant results. Summing up, he writes: "These findings among other things raise doubts whether the big, monopolistic, conglomerate corporation is as efficient an engine of technological change as disciples of Schumpeter (including myself) have supposed it to be. Perhaps a bevy of fact-mechanics can still rescue the Schumpeterian engine from disgrace, but at present the outlook seems pessimistic" (1122).

In fact Scherer's pessimism was to be borne out. Over the next few decades fact-mechanics (or econometricians as they are now called) did not find evidence in favor of the Schumpeterian model; in fact, quite the opposite was the case, as the empirical tide turned against Schumpeter.

1.2.2 Methodological Challenges

Before skipping ahead to the recent microeconomic literature, our explorer would do well to consider some of the methodological difficulties faced by empirical researchers in this area. The early literature failed to reach robust conclusions principally because of a number of difficult methodological problems that were not dealt with, in large part due to lack of data.[10]

First, it turned out to be important to control for other firm and industry characteristics that affect innovation. This is because these other characteristics are correlated with firm size and market structure. For example, if we showed that firm size was positively associated with innovative output, but we had not controlled for firm age, then it could be the case that firm size is correlated with age (e.g., firms get bigger as they get older) and that firm age is also correlated with innovate output, and that this led to a spurious correlation between firm size and innovation. Unless we control for at least the main observable and unobservable characteristics, we can not be sure that we are really picking up the relationship between size and innovation.

Second, there is a problem of reverse causality. While firm size or market structure is likely to affect innovation, it is also the case that successful innovation affects market structure. Firms that are successful innovators will either have lower costs (so be able to sell at a lower price) or have superior quality goods, and in either case will gain market share.[11] To help deal with these first two difficulties, it is important to have panel data—repeated observations of the same individuals over time. Panel data in itself does not solve these problems. What is important is that there is exogenous variation in the degree of competition, for example, policy changes that make entry easier or less costly. If we are willing to assume that many of the firm characteristics that are correlated with market power are constant over time, then firm fixed effects can be used to control for them.[12] In addition, if we are willing to assume that market structure is predetermined (i.e., that feedback from innovation to market structure only affects future market shares, and the anticipation of innovation does not affect current market structure), then repeated observations of the same firm allow us to use lags of market structure.

Third, the relationship we are interested in is between product market competition and innovation, while the early literature largely focused on the relationship between firm size or market concentration and innovation. These may not be good measures of the degree of competition, and may in fact reflect other differences, for example, a firm's ability to access finance. Boone (2000) shows it is not always the case that an increase in competition reduces firm size, price cost margins, or concentration.[13] It can be difficult to obtain good measures of the degree of product market competition in an industry, and recent work has paid careful attention to this. Recent work has used a measure of rents, or the Lerner Index (see, e.g., Nickell 1996). This is the measure implied by much of the theory used in what follows and has several advantages over indicators such as market shares or a Herfindahl or concentration index. In order to measure any of those, it is necessary to have a definition of both the geographic and product boundaries of the market in which the firm operates. This is particularly important in applications where firms operate in international markets, so that traditional market concentration measures could be extremely misleading.[14]

A related, but somewhat separate point, is that it may be the case that firms with greater market share are more innovative (than those with lower market share), but if we are interested in aggregate (or industry) innovations then we have to weigh this against the fact that there will be fewer firms innovating in more concentrated markets, and thus industry innovation could go up or down.

Fourth, it can be difficult to measure innovative output. Measures that are commonly used at the firm level are R&D spending, patenting activity, innovation counts, and total factor productivity (TFP). However, each of these has its problems. R&D expenditure is an input, not an output. In addition, in many countries or time periods it has not been mandatory for firms to report it. For example, in the United Kingdom prior to 1990, it is frequently not reported. Patenting activity, innovation counts, and TFP are all output measures, but each has its own problems. Patents are a very heterogeneous measure of innovation.[15] One patent can represent a path-breaking new technology worth billions of dollars, while another can represent a fairly incremental improvement in an existing technology worth only tens of thousands of dollars. In order to get around this problem, many researchers use citation-weighted patents[16] or use stock market data to assign a value to a patent. Another problem with patents is that the propensity to patent, and the degree to which they provide protection of intellectual property rights, varies substantially from industry to industry. For example, patents are widely used in the pharmaceuticals industry, but rarely used in the computer software industry. Innovation counts have also been used. An example is the Science Policy Research Unit (SPRU) innovations dataset for the United Kingdom.[17] These type of data are laborious to collect and suffer from problems similar to those of patents in the sense that they are very heterogeneous in their value and it is difficult to obtain consistent measures, particularly over time. However, where available, they can provide a rich source of detailed information. The final measure, TFP, is a measure of technological progress (and thus implemented innovative activity), but it can be difficult to accurately measure because of the well-known problem that commonly used measures of TFP are themselves biased in the presence of imperfectly competitive product markets.[18]

After considering these difficulties, our intrepid explorer may well consider giving up and going home. However, a combination of improved data availability (and, in particular, the availability of firm-level panel data sets), better econometric methods and more computing power meant that many of these problems could be tackled by the mid-1990s.

1.2.3 The 1990s

Rather than survey all of the papers in this area, we have steered our explorer to two specific papers that directly addressed these issues—Nickell 1996 and Blundell, Griffith, and Van Reenen 1999.[19] These both use data on firms listed on the London Stock Exchange. The United Kingdom turned out to be a good place to study the relation between product market competition and innovation because there have been a large number of policy changes that led to (relatively) exogenous variation in the nature and magnitude of product market structures and competition. These included the large-scale privatizations of the 1980s and 1990s, reforms associated with EU integration, and the opening up of markets in numerous other ways.

Nickell (1996) Nickell (1996) considered the link between market structure and both the level and growth rate in TFP.[20] Nickell's paper was the first to tackle these empirical issues head on. Using firm-level panel data, he was able to control for unobservable (correlated) characteristics that were constant over time. He developed and used better measures of product market competition including market share, concentration, import penetration, rents, and survey-based measures. The measure based on rents is particularly interesting as it is robust to a number of concerns about identifying the markets in which firms operate. Primarily, it does not require the econometrician to be able to define and observe the entire market(s) in which the firm operates, unlike alternative measures such as market share, Herfindahl, or concentration indices.

While Nickell's measure of TFP suffers from the bias mentioned previously (it is negatively correlated with the degree of product market competition), Nickell points out that this would only work against his findings—it would make it more likely to find evidence *supporting* the Schumpeterian hypothesis of competition being bad for innovation. Nickell provides convincing support for the idea that tougher competition in the product market is associated with higher growth rates in TFP—higher concentration and a higher level of rents are associated with lower growth rates of TFP.

Nickell estimates an augmented production function. The basic equation of interest (Nickell 1996, eq. (5)) is

$$\ln \frac{Y_{it}}{K_{it}} = \phi_1 \ln\left(\frac{Y_{it-1}}{K_{it}}\right) + \phi_2 \ln\left(\frac{L_{it}}{K_{it}}\right) + \phi_3 MS_{it-2} + t(\phi_4 SZ_i + \phi_5 RT_i + \phi_6 C_j + \phi_7 IMP_j) + X_{it}^{'}\lambda + e_{it}, \tag{1.5}$$

where i indexes firms, Y is output, L is labor inputs, K is capital inputs, MS is firm market share, t is a time trend, SZ is firm size, RT is the level of rents earned by the firm (normalized by value-added), C is an industry-level concentration ratio (market share of the top five firms), IMP is industry import penetration, and X is a vector of other variables including firm and time effects (to capture unobservable characteristics of the firm that may be correlated with the variables of interest).

Notice that this specification implies that market share, in contrast to other competition measures in equation (1.5), affects the firm's *level* of TFP. The other measures (SZ_i, RT_i, C_j, IMP_j) are measured at the cross-sectional level (they are not time-varying) and are multiplied by t, which represents a time trend. This means that they are modeled as affecting the growth rate of TFP. When equation (1.5) is differenced, to remove firm-level unobservable characteristics, the model becomes

$$\Delta \ln \frac{Y_{it}}{K_{it}} = \phi_1 \Delta \ln\left(\frac{Y_{it-1}}{K_{it}}\right) + \phi_2 \Delta \ln\left(\frac{L_{it}}{K_{it}}\right) + \phi_3 \Delta MS_{it-2} + \phi_4 SZ_i + \phi_5 RT_i + \phi_6 C_j + \phi_7 IMP_j + \Delta X_{it}^{\prime}\lambda + \Delta e_{it},$$

where Δ represents the one-year difference (e.g., $\Delta x_{it} = x_{it} - x_{it-1}$). This makes it clear that it is the levels of SZ_i, RT_i, C_j, and IMP_j that affect the growth rate of TFP. Our interpretation of the difference between an effect on the level of innovation versus the growth rate is something we return to in the next chapter. For now it suffices to note that the main results of interest to us here in Nickell's (1996) paper are those on the growth rate of TFP.

The estimates of the coefficients of interest in Nickell's basic specification all indicate that increased competition is associated with a higher level and faster growth rates of TFP. Column (1) of table 1 in Nickell 1996 suggests that the coefficient on market share (ϕ_3) was -3.49 (and was statistically significant with a t-statistic of 2.1), suggesting that firms with lower market share had higher *levels* of TFP. The coefficients on size (ϕ_4) and imports (ϕ_7) were positive, but not statistically significant. The coefficient on rents (ϕ_5), which is a measure of the degree of competition the firm faces in the product market (the higher a firm's rents, the less competitive the market), is -0.13 (and was statistically significant with a t-statistic of 2.9), suggesting that firms in more competitive markets had higher growth rates of TFP.

What economic interpretation do we put on these estimates? Nickell reports the distribution of rents as being zero at the twentieth percentile and 0.29 at the eightieth percentile. Looking at a very similar data set, we can see that the mean and median are around 0.20. Combining this with the estimated coefficient on RT_i of -0.13, we get that increasing competition by going from the eightieth percentile in the distribution of rents to the twentieth (reducing rents means increasing competition) increases TFP growth by around 3.8 percentage points. This is a large and economically significant effect. Moving from the median to the twentieth percentile would be associated with an increase in TFP growth of around 1.2 percentage points.

The impact of competition on innovation, as measured by TFP, is illustrated in figure 1.3. This figure plots the values of $-0.13{*}RT_i$, normalized to zero at the lowest level of competition. The far right hand of the x-axis (where $(1 - RT) = 1$) represents perfect competition (zero rents), while the far left-hand end of the x-axis represents a low level of competition (rents of 30 percent).

The coefficient on industry concentration (ϕ_6) was -0.12 (and was statistically significant with a t-statistic of 2.1). Performing a similar exercise, we find that the industry concentration ratio at the twentieth percentile is 0.25 (the top five firms make up 25 percent of total output) and at the eightieth is 0.60 (the top five firms account for 60 percent of output), with a mean of 0.41 and median of 0.43. If an industry moves

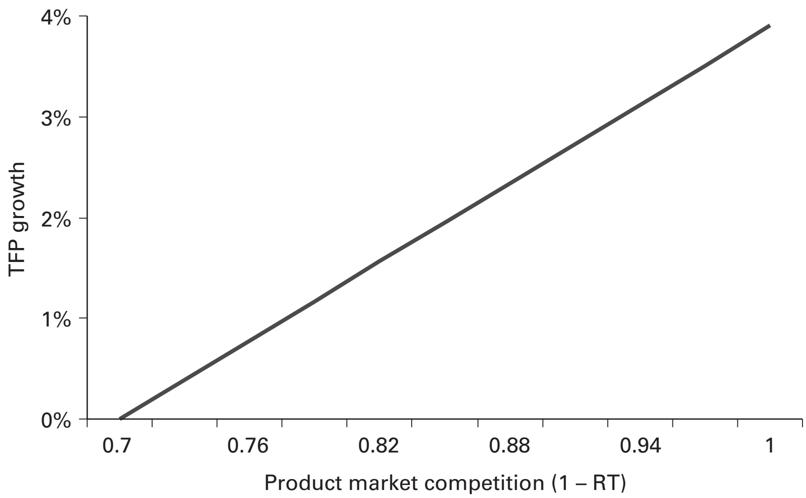

Figure 1.3
Relationship between product market competition and total factor productivity implied by Nickell 1996 results

from being very concentrated (at the eightieth percentile) to being fairly unconcentrated (at the twentieth percentile), ceteris paribus, this will be associated with an increase in TFP growth of 4.2 percentage points. Again, this is a large impact.

Another way to look at the economic importance of these estimates is to look at how much of TFP growth is explained by differences in competition. Nickell creates an index of competition that equals

$$\phi_5 RT_i + \phi_6 C_j. \tag{1.6}$$

Nickell reports the value of this across industries and they are substantial, as shown in table 1.1.[21] This shows the differences that arise in average industry growth rates due to differences in the level of competition across these industries, holding everything else constant. For example, TFP growth in electrical engineering was 2.4 percentage points lower, on average, due to low levels of competition, while mechanical engineering experienced TFP growth that was around 1 percentage point higher due to relatively higher levels of competition.

Blundell, Griffith, and Van Reenen (1999) Another micro study by Blundell, Griffith, and Van Reenen (1999), also uses U.K. firm-level panel data, but rather than using TFP it uses the SPRU innovation

Table 1.1
Percentage point TFP growth rate differentials generated by differences in competition

Food, drink, and tobacco	−.2	Metal goods (other)	.8
Chemicals	−.8	Textiles	.9
Metal manufacture	−1.7	Clothing and footwear	1.0
Mechanical engineering	1.0	Bricks, pottery, and glass	2.0
Instrument engineering	−.6	Timber and furniture	1.6
Electrical engineering	−2.4	Paper, printing, and publishing	1.9
Vehicles	−1.3	Other manufacturing	−2.2

Source: Nickell 1996, table 3.
Note: These are differentials from the unweighted mean.

count to measure innovative output.[22] Competition is measured using market share, concentration, and import penetration. The main contribution of this paper was to embed the empirical work in a clear theoretical framework in order to address the question of *why* market dominance enables firms to be more innovative. There were two main interpretations of Schumpeter's work emphasized in the literature. The first said that financial market failures meant that firms had to rely on their own internal sources of funds in order to finance innovation, and that larger firms had deeper pockets and were thus better able to do this. The second focused on the incentive effects of market power as highlighted, for example, by Gilbert and Newbery (1982), who argued that monopolists will tend to innovate more than entrants because of the reduction in total industry profits that the incumbent suffers due to entry. In contradiction to this is the displacement effect emphasized by, for example, Reinganum (1983), whereby a monopolist is less keen on innovating because this will replace some of the current stream of rents, while the entrant does not suffer from such disincentives.

Blundell, Griffith, and Van Reenen (1999) also tackle a number of the econometric issues, incorporating dynamics and controlling for firm fixed effects in a nonlinear model. The main equations estimated are an innovation equation

$$I_{it} = \exp(\alpha_1 MS_{it-1} + \alpha_2 SZ_{it-1} + \alpha_3 C_{jt-1} + \alpha_4 IMP_{jt-1} + X'_{it}\lambda + u_{it}) \qquad (1.7)$$

and a market value equation

$$\ln V_{it} = \alpha_5 \ln K_{it} + \alpha_6 G/K_{it} + \alpha_7 MS_{it} + \alpha_8 (G/K_{it}) MS_{it} + X'_{it}\lambda' + \varepsilon_{it}, \qquad (1.8)$$

where I is a count of innovation from the SPRU dataset, V is firm market value on the London Stock Exchange, and G is stock of innovations; as in the previous case, i indexes firm, MS is firm market share, SZ is firm size, C is an industry-level concentration ratio (market share of the top five firms), IMP is industry import penetration, K is tangible capital, and X is a vector of other variables including firm and time effects.

Blundell, Griffith, and Van Reenen's results showed that less competitive industries (those with higher concentration levels and lower imports) had fewer aggregate innovation, as shown in table 1.2. Column 1 shows the estimates from a regression of the form shown in equation (2.7), while columns (2)–(6) extend this to alternative dynamic specifications. Focusing on columns (2) and (3), which represent alternative dynamic specifications, we see that the coefficient on the concentration ratio (α_3) is negative and significant and the coefficient on market share (α_1) is positive and significant. This suggests that tougher competition (a lower concentration ratio) is associated with higher levels of innovation, even though within industries it was the higher market share firms that innovated most frequently.

While the point estimates on the market share coefficient differ across columns (2) and (3) the economic interpretation is also not the same, because of the different dynamic specifications. In fact, the two estimates suggest very similar short-run impacts and elasticities. The short-run impact of market share on innovation is 0.15 in column (2) and 0.16 in column (3), and the elasticity of innovation with respect to market share, evaluated at the mean market share, is 0.08 in column (2) and 0.10 in column (3).

One of the contributions of Blundell, Griffith, and Van Reenen is that considering, on the one hand, the relationship between market structure and innovation, and, on the other hand, the impact that market structure has on the relationship between innovation and market value enables the authors to distinguish between the two reasons for seeing a positive correlation between market share and innovation—financial constraints or incentive effects. In column (4) of table 1.2, Blundell, Griffith, and Van Reenen included a measure of free cash flow and showed that market share was not simply picking up the effect of greater liquidity in larger firms. Estimates of equation (1.8), shown in table 1.3, examine this issue further. The econometric results shown there lend support to the Gilbert and Newbery preemption effect (1982) discussed in section 1.1.2. The coefficient on the interaction between market share and firms' knowledge capital stock (α_8) was posi-

Table 1.2
The innovation equation estimates

	(1)	(2)	(3)	(4)	(5)	(6)
G_{-1}	—	—	0.123	0.122	0.156	−0.086
			0.052	*0.052*	*0.037*	*0.352*
$\ln(G_{-1})$	—	0.331	—	—	—	—
		0.081				
G_{-1} dum	—	0.540	—	—	—	—
		0.750				
MS_{-1}	4.318	1.336	2.534	2.568	3.207	3.739
	0.988	*0.451*	*0.713*	*0.699*	*1.028*	*3.278*
Conc_{-1}	−1.967	−1.498	−2.198	−2.190	−1.759	−6.499
	0.936	*0.676*	*0.976*	*0.896*	*1.135*	*11.111*
Imports_{-1}	1.214	0.987	1.258	1.316	1.597	0.841
	0.925	*0.806*	*1.118*	*1.312*	*1.254*	*2.941*
K_{-1}	0.894	0.124	0.208	0.200	0.060	0.036
	0.228	*0.122*	*0.181*	*0.191*	*0.244*	*0.494*
Cash_{-1}	—	—	—	−0.207	—	—
				0.534		
$G\text{-Prod}_{-1}$	−0.282	−0.466	−0.422	−0.416	−0.133	−0.768
	0.567	*0.384*	*0.547*	*0.548*	*0.629*	*3.612*
$G\text{-User}_{-1}$	4.917	2.562	3.278	3.288	2.299	1.662
	1.740	*1.381*	*1.900*	*1.920*	*2.174*	*2.596*
$\ln(G_0)$	—	0.452	0.838	0.829	0.862	—
		0.106	*0.114*	*0.114*	*0.129*	
G_0 dum	—	0.696	1.825	1.660	2.062	—
		0.793	*0.739*	*0.750*	*0.862*	
1973–1974	−0.300	−0.432	−0.957	−0.926	—	—
	0.153	*0.162*	*0.494*	*0.478*		
1980–1982	−0.993	−0.676	−0.934	−0.959	—	—
	0.209	*0.252*	*0.616*	*0.614*		
Constant	−2.956	−0.195	−0.327	−0.326	—	—
	0.588	*0.408*	*0.689*	*0.698*		
Time dummies	no	no	no	no	yes	yes
Observations	3511	3511	3511	3511	3511	3211
Time period	1972–1982	1972–1982	1972–1982	1972–1982	1972–1982	1972–1981
v_1	1.210	−0.270	−0.474	−0.472	−0.059	
v_2	2.980	0.271	−0.573	−0.572	−0.913	

Source: Blundell, Griffith, and Van Reenen 1999, table 4.1.
Notes: Standard errors are in italics and allow for general heteroskedasticity and autocorrelation. Dummy variables for GEC and ICI are included in all columns except (6). In columns (2)–(5) instruments include a single lag of each variable and the initial value of firm level variables (MS and K). v_1 and v_2 are the standard serial correlation statistics from Arellano and Bond (1991) distributed $N(0,1)$ under the null of no serial correlation. In column (6) instruments are lags of all variables.

Table 1.3
Market value in levels

	$\ln(V)$ (1)	$\ln(V)$ (2)	$\ln(V)$ (3)	$\ln(V)$ (4)	$\ln(V/K)$ (5)	$\ln(V/K)$ (6)	$\ln(V/K)$ (7)
$\ln(K)$	0.688	1.098	1.095	1.088	—	—	—
	0.016	*0.061*	*0.061*	*0.058*			
G/K	1.928	2.064	1.421	1.615	1.582	1.840	4.370
	0.933	*0.970*	*0.947*	*0.923*	*0.921*	*0.885*	*1.852*
MS	0.600	0.341	0.075	0.007	0.068	0.277	−0.001
	0.144	*0.380*	*0.350*	*0.329*	*0.327*	*0.327*	*0.671*
$MS * (G/K)$[a]	—	—	1.588	1.733	1.745	1.767	1.715
			0.404	*0.402*	*0.403*	*0.376*	*0.952*
Conc	—	—	—	0.374	0.379	0.385	0.441
				0.082	*0.082*	*0.082*	*0.211*
Imports	—	—	—	−0.316	−0.319	−0.315	−0.321
				0.099	*0.099*	*0.099*	*0.265*
Union	—	—	—	−0.260	−0.259	−0.264	−0.267
				0.113	*0.113*	*0.113*	*0.271*
G-User	—	—	—	−0.345	−0.331	−0.357	−0.543
				0.222	*0.223*	*0.223*	*0.500*
G-Prod	—	—	—	0.255	0.252	0.260	0.331
				0.062	*0.062*	*0.062*	*0.136*
$\ln(V_0)$	—	0.151	0.152	0.141	0.141	0.142	0.144
		0.014	*0.014*	*0.014*	*0.014*	*0.014*	*0.036*
G_0	—	0.145	0.004	−0.030	−0.029	−0.030	−0.026
		0.040	*0.049*	*0.056*	*0.048*	*0.047*	*0.120*
G_0 dum	—	−0.182	−0.184	−0.181	−0.179	−0.184	−0.214
		0.029	*0.028*	*0.028*	*0.028*	*0.028*	*0.069*

$\ln(K_0)$	—	−0.413	−0.406	−0.408	−0.333	−0.335	−0.331
		0.059	*0.059*	*0.057*	*0.015*	*0.015*	*0.041*
MS_0	—	−0.491	−0.264	−0.031	−0.089	−0.305	−0.033
		0.408	*0.365*	*0.343*	*0.342*	*0.342*	*0.720*
Sargan (df)	—	—	—	—	—	325 (220)	207 (192)
p-value	—	—	—	—	—	0.000	0.216
Observations	3511	3511	3511	3511	3511	3511	3211
Years	1972–1982	1972–1982	1972–1982	1972–1982	1972–1982	1972–1982	1973–1982
v_1	10.21	11.03	10.98	11.16	11.11	11.00	9.82
v_2	9.36	10.19	10.13	10.27	10.19	9.97	8.16
$R^2(P-S)$	0.842	0.862	0.863	0.863	0.389		

Source: Blundell, Griffith, and Van Reenen 1999, table 4.2.

Notes: A full set of time dummies, a dummy for the chemical sector, ln (inventories) are included as additional controls in all specifications. Initial stock of log (inventories) is included in columns (2)–(7). Standard errors are in italics and allow for arbitrary heteroskedasticity. All industry-level variables are assumed exogenous. Instruments for firm-level variables (X_i) are columns X_{t-1} in (1)–(5), X_{t-1} to X_{t-7} in column (6), and ΔX_{t-1} to ΔX_{t-7} (except the interaction) in column (7). The Sargan test is distributed χ^2 under the null of instrument validity. $R^2(P-S)$ are calculated as in Pesaran and Smith 1995.

[a] Coefficients and standard errors divided by 100.

tive and statistically significant. This suggests that higher market share firms get a bigger payoff from an innovation, giving them a greater incentive to preemptively innovate.

In a slightly different form of the same specification, Blundell, Griffith, and Van Reenen (1999) estimate that an innovation is worth on average around £2m. This estimate is in line with others in the literature; for example, Geroski, Machin, and Van Reenen (1993) estimate a similar impact using data on profitability.

1.3 Conclusion

Not only the informal thinking of many economists, but also the empirical evidence of the mid-nineties, seemed to contradict the basic theoretical prediction that product market competition is detrimental to innovation and growth. As Nickell (1996) summarizes, "this general belief in the efficacy of competition exists despite the fact that it is not supported either by any strong theoretical foundations or by a large corpus of hard empirical evidence in its favor" (725).

So the theoretical work and empirical evidence were at odds. The empirical literature suggested that more competitive market structures were associated with greater innovative output, an idea that had much support in policy circles. However, the empirical models were missing something—in particular, work so far (excepting Scherer's early work) had only looked for linear effects. Our knowledge that, for example, patent protection (the granting of a time-limited monopoly to a firm) was good for innovation suggested that, at least over some ranges, less competition could be conducive to innovation. But this was not being captured in the empirical work.

There was also a need to reconsider the theoretical models that suggested that more competitive market structures had an unambiguously negative effect on innovation and productivity growth. These models were missing something too, including the possibility that innovations, particularly vertical innovations, could be made by incumbent firms in order to preempt or escape competition and entry.

In the following chapters, our explorer digs into these cracks in the theories and empirical studies on competition and growth and finds that harmony can be restored. Before doing that, however, we return in the next chapter to the common wisdom that competition increases productive efficiency.