CHAPTER 1

Profits and the process of competition

DENNIS C. MUELLER

I Two views of competition

Two views about competition exist. The first sees competition as a process for allocating resources to their optimal uses. The price mechanism is the instrument for achieving this goal, and when it functions properly, equilibria emerge with prices equated to marginal social costs of production. When it malfunctions, equilibria exist with some prices above marginal costs, and society suffers a welfare loss from the underconsumption of these goods. Such malfunctions are usually attributed to an insufficient number of buyers or sellers. Monopoly is seen as the antithesis of competition. Thus, under the first view, competition is seen as a process for determining prices and quantities, the allocation of resources for a given set of tastes and technological opportunities. At its zenith, competition produces an equilibrium set of prices that induce a Pareto optimal allocation of the economy's goods and services. Such equilibria are anticipated so long as monopolistic elements are absent.

The other view of competition sees it not as a process for allocating a given stock of resources but as a process for transforming these resources into new products and production techniques. Competition takes the form not of lower prices for an existing set of products but of new and improved ideas, and these in turn are the property of the individual(s) who created them and his/her/their employer. In the first instance, competition for a new product is competition for a newly created monopoly. With time the monopoly disappears as other firms imitate and improve upon the new product. Thus, monopoly is an integral part of a dynamically competitive process, a passing stage in an industry's evolution whose presence might signify progressive good performance just as readily as poor performance.

When competition is viewed as a dynamic process of new product and process creation, the concept of equilibrium does not play an important role. What is of interest is not the constellation of prices and allocation of resources at a particular point in time but their movements over time. The

perspective is that of a system in flux, of constant disequilibria evolving through time, rather than of a system in a state of equilibrium at a particular point in time.

It is the first view of competition that informs most economic analysis and underlies most model building. And it is the first view of competition that is predominant in the economics classroom. Yet it is arguably the second view that more accurately describes actual competition. The essays in this volume attempt to model and test this second view of the competitive process.

II The static model

The origins of the first view of competition can be traced back to Adam Smith (1776) or more directly to Augustin Cournot (1838). From Cournot we obtain a clear definition of a market equilibrium and a prediction that price and number of sellers are inversely related at the equilibrium. The Cournot model might be regarded as the theoretical foundation for a generation's empirical research in industrial organization relating firm or industry profitability to industry concentration levels.

This research established fairly conclusively that profitability and concentration are positively correlated, as the Cournot model or a small numbers-collusion hypothesis predicts (Weiss, 1974). Given the assumptions underlying the Cournot-collusion hypothesis, the inference must be drawn that the divergence between price and costs is greater in concentrated industries, and from this it follows that welfare losses must be greater in more concentrated industries.

The assumptions underlying these inferences leading to a positive correlation between profit and concentration have been challenged, however. Yale Brozen (1971a, b) argued that firms with efficiency advantages over their competitors tended to grow faster than their competitors. Therefore, it would be these more efficient firms that would grow to fill the ranks of the largest firms in an industry. At any point in time, the most concentrated industries in an economy would contain many that had become so through recent, rapid growth due to efficiency advantages. Brozen argued that the normal process of competition and imitation would eliminate the efficiency advantages of these firms, reducing both industry profitability and concentration. Brozen thus reasoned that the positive association between profit rates and industry concentration levels was a transitory, disequilibrium phenomenon. He presented empirical evidence indicating that the positive association between these variables that Bain (1956) had observed did disappear in a few years.

Brozen's critique of Bain's early results and of the generation of research it spawned raises two fundamental objections to the underlying

assumptions of this model. The first questions whether the costs and prices of all firms within an industry can be assumed to be the same. The second questions whether the relationship between a firm or industry's profitability at a given point in time and other firm–industry characteristics can be assumed to represent a stable, long run equilibrium relationship. Both objections indirectly question the applicability of the first view of the competitive process to the study of company profitability. To investigate efficiency differences across firms, which change over time, one must adopt a dynamic view of the competitive process.

III The Schumpeterian perspective

The second representation of the competitive process owes its origin to Joseph Schumpeter (1934, 1950), of course. The salient feature of Schumpeter's description of the capitalist process is its dynamic nature. Iconoclastic entrepreneurs introduce innovations – new products, new production processes, new marketing techniques, new organizational structures – that create temporary monopolistic advantages over their competitors. These transitory monopolies create pockets of profits, which in turn provide the incentive for imitators to step forward and thereby drive these profits back to zero. Thus, the "process of creative destruction" proceeds: innovation creating monopoly, monopoly creating profits, profits creating imitators until a state of normalcy returns, only to be followed by new innovations and a repeat of the cycle.

This Schumpeterian image of the competitive process often seems to underlie more informal arguments by economists in favor of capitalist institutions over socialist institutions, unregulated markets over regulated markets, or an explanation for why some developed countries outperform others over the long run. Yet the Schumpeterian perspective has had little impact on the development of more formal models of market behavior. Efforts to move the profession in this direction (e.g., Clark, 1961; McNulty, 1968; Kirzner, 1973), although often applauded, do not launch a subsequent stream of research that develops their initial insights. The Schumpeterian perspective remains just that, a perspective on the nature of competition rather than a model of the competitive process.

The most important exception to this generalization is perhaps the evolutionary model of capitalism of Nelson and Winter (1982).[1] Nelson and

[1] A second important exception might appear to be the contestable market theory of Baumol, Panzer, and Willig (1982). Entry and exit are the heart of Schumpeter's process of creative destructive, just as the conditions of entry and exit are central to defining the contestability of a market. But in contrast to Schumpeter's dynamic depiction of the competitive process, the contestable market theory is entirely static. It relates *conditions*

Winter consciously eschew standard neoclassical models of profit-maximizing firms operating in competitive markets and the notion that these markets are in equilibrium. Rather, they trace the evolution of firms and industries over time using simulation techniques. Although these simulations trace out a rich mosaic of a capitalist economy's evolution, a mosaic that accords well with actual experience in many ways, they do not constitute a formal test of a Schumpeterian or evolutionary model against, say, a neoclassical model. They raise rather than answer the question of how to test a Schumpeterian model of the competitive process.

Such a test would seem to have to consist of at least two elements: first, an examination of the histories of various industries to see whether they follow the innovation–imitation–maturity cycle sketched by Schumpeter and, second, an examination of the profit histories of firms and industries to see whether the process of dynamic competition does indeed erode abnormal profits over time.

Case study evidence on product and industry life cycles seems to support a Schumpeterian description of an industry's evolution (see, e.g., Gort and Klepper, 1982; Klepper and Graddy, 1984). Initial innovations are followed by the entry of numerous "imitators," which in turn leads to a shake-out phase in an industry's evolution, which might be likened to Schumpeter's gale of creative destruction. Whereas this general pattern accords with the Schumpeterian view of industry dynamics, industry life cycles appear to enfold at substantially different speeds, and exceptions do exist (e.g., Gort and Klepper, 1982; Klepper and Graddy, 1984). Moreover, the studies of product life cycles have not investigated the profit performances of the innovating and imitating firms. Indeed, very little is known about the intertemporal patterns of profitability for individual firms and industries in general. This void will be filled to some extent by the essays in this volume.

We seek to test two main tenets of the Schumpeterian thesis: (1) that the competitive process does successfully erode positions of excess profits, that is, in the long run all economic rents tend toward zero, and (2) that the erosion process proceeds quickly. Although the first hypothesis can be clearly formulated and tested – rents are either zero or they are not – the second hypothesis is more equivocal. How quick is quick? We shall not actually try to answer this question but rather present evidence of the speed of adjustment of profits to their long run equilibrium values.

of entry to *levels* of profit in equilibrium. The dynamics of attaining equilibrium are not addressed. Testing the contestable market theory has proceeded in the tradition of other empirical work in industrial organization by relating profit rates to market structures in cross-sectional analyses.

IV Previous research

Yale Brozen's (1971a, b) critiques of Joe Bain's work were the first studies to raise, if only indirectly, the question of the intertemporal pattern of profitability and its relationship to market structure characteristics. David Qualls (1974) responded to Brozen's attack on Bain by pointing out that high concentration was a necessary but not sufficient condition for persistent above normal profits. Where entry barriers are low, high profits attract new entrants, which in turn drive profit rates down. Thus, Qualls argued that one should expect a stable concentration–profits relationship only in industries with high entry barriers. He presented evidence of a continuing positive association between profits and concentration for those firms and industries having high entry barriers.

Although the profits–concentration relationship remained significant after a decade, Qualls's estimates indicated a weakening of the relationship and thus were consistent with an eventual disappearance of all above normal profits. Indirectly, his results posed the salient questions of this study. Do profits converge on a common, normal rate of return, and how is this convergence process related to firm and industry characteristics?

A large literature dating back a half century has been concerned with the intertemporal variability of profit rates in response to business cycle factors. These studies have predicted changes in profit rates or price–cost margins over the business cycle, with these changes being related to industry characteristics like concentration and capital intensity. In general, the studies have assumed that price–cost margins change with varying business cycle conditions due to changes in the abilities of oligopolists to coordinate behavior (e.g., changes in conjectural variations; see Qualls, 1979). They have not been concerned with the long run equilibrium values of firm and industry profit rates and the adjustment process to this equilibrium as a result of the entry and exit of firms and intraindustry shifts of resources – the concerns of the present set of studies. Thus, we do not attempt to build off of this previous research, but rather we adjust the data to mitigate the influence of business cycle factors on our results.

The first direct test of the hypothesis that individual profit rates converge on a common normal value was presented by myself (Mueller, 1977). Using data for 472 firms from 1949 to 1973, I rejected the null hypothesis of convergence on a common rate of return. Subsequent research on a sample of 551 companies confirmed this earlier conclusion and presented evidence on the relative importance of various firm and industry characteristics in explaining permanent differences in company profit rates (Mueller, 1986).

Connolly and Schwartz (1985) raised several objections to the methodology of my 1977 study and presented alternative tests of the competitive environment hypothesis. Their evidence supported the hypothesis for below normal profit firms. In the long run, the profits of companies that are at one time below the normal level do tend to rise toward that level. But for a substantial fraction of firms with above normal returns, the process is incomplete. These firms appear to earn above normal rates of return indefinitely.

David Levy (1987) estimated the adjustment rates to long run equilibrium values for profit rates measured at the industry level using U.S. data and a model similar to that used in this study. He estimated very slow adjustment rates when he did not include separate industry intercepts and fairly rapid adjustment rates when he did include them. His results at the industry level are similar to those reported by Kessides in Chapter 4 and to those of the other studies at the firm level. Significant differences in permanent profit rates exist across industries, but adjustments to these permanent rates are rather quick (less than five years in Levy's results).

Geroski and Jacquemin (1988) present estimates of permanent rents and the speeds of adjustment to long run equilibrium for France, West Germany, and the United Kingdom. Their findings, based on a smaller sample of firms, are somewhat at variance with those reported in this volume for these three countries. Geroski and Jacquemin find a greater degree of profits persistence in the United Kingdom than in France and West Germany, whereas the results in this volume indicate that permanent rents are higher in France, with West Germany and the United Kingdom being relatively similar.

The first test of whether persistent differences in profit rates exist across firms for a country other than the United States was the Odagiri and Yamawaki (1986) study of Japan. They found persistent differences in profit rates across Japanese companies and that a significant component of these differences can be explained by industry characteristics.

V The present, international comparison

Regardless of which view of the competitive process one takes, the question of whether profit rates converge on a common, normal return and the measurement of the speed of convergence are central to both a positive and normative discussion of economic competition. If convergence is quick and complete, then competition among existing firms within an industry, the entry and exit of firms, and the threatened entry of firms must be presumed to be sufficiently swift and strong to bring firm and industry profit rates quickly back to their long run competitive equilibrium

levels. Although deviations from this equilibrium at any point in time exist, the competitive process can be seen as functioning rapidly enough so that substantial and permanent welfare losses from price–cost deviations do not exist. Similarly, if one thinks in terms of dynamic competition, the entry of imitators and exit of failed firms must be presumed to proceed rapidly enough so that even innovators' profits are relatively short-lived.

The existence of permanent rents and/or a slow adjustment to long run competitive equilibrium values raises some troubling questions, however. Measured in static terms, the cumulative welfare losses from deviations between prices and costs could be large. The knotty question then must be faced as to whether these welfare losses from allocative inefficiency are somehow offset by the welfare gains from innovations that create long-lived monopoly positions.

Although industrial organization economists have certainly been aware of the importance of these dynamic competition issues, this awareness has generally not informed empirical studies of firm and industry profitability. The existing literature consists of but a few studies of the movements of profits over time, and these employ disparate methodologies.

The present volume seeks to help fill this void by presenting a set of studies of intertemporal patterns of company profitability in different countries, all of which build from a common methodological foundation. This foundation is developed in the next chapter. It is followed by separate studies for the United States at the firm and industry level. These in turn are followed by individual chapters of profitability at the company level for Canada, the Federal Republic of Germany, France, Japan, and the United Kingdom. The original plan was to include also a chapter on Sweden, but it was not completed in time for inclusion in the final volume. Preliminary results were supplied by Gunnar Eliasson and are included in the overview, Chapters 10 and 11, however. These two chapters summarize the results for the individual countries and draw some implications from them.

The countries studied differ in size, progressivity, and international competitiveness. With respect to size, they range from Canada up to the United States. The United States for the period studied (1950–72) and Japan are relatively closed economies. Others were quite open to international competition. Japan has been the world's pacesetter in productivity, income, and export growth. The United Kingdom has been at the other pole, epitomizing slow growth and economic retardation.

These contrasts raise several interesting questions concerning the intertemporal profit patterns of the different countries. Are permanent profit differences greater and adjustments to equilibrium slower in the relatively

closed economies of Canada, Japan, and the United States than in the other countries? Has Japan's spectacular performance in international markets been led by a group of dynamically progressive firms, which have been able to translate this success into persistently above normal profits? Is the sluggish overall performance of the United Kingdom evident in individual company profit rate patterns, for example, in relatively slow adjustment speeds to long run equilibrium levels? These and other questions are addressed in this volume.

Almost all of the analyses in this book are based on rates of return on corporate assets. These accounting measures of performance have been subject to substantial criticism in recent years, and their usefulness has been the topic of a hot debate. We do not wish to join this debate here but feel some discussion of the issue is warranted. The appendix of this chapter discusses the major questions. Those willing to accept accounting measures of profitability as legitimate measures of performance can move directly to Chapter 2.

Note to the reader: Chapters 3–9 make use of the methodology of Chapter 2. Each is written more or less as a self-contained entity, however, so that following Chapter 2, the reader can read from the ensuing seven chapters in any order and number(s) he or she chooses. Chapters 10 and 11 overview the results and conclusions of the book.

Appendix

Accounting returns versus economic returns

All of the studies in this book use accounting profits and assets data. Although accounting data have various well-known deficiencies, it has generally been felt that they can be used with caution to analyze company performance. Recent papers by Fisher and McGowan (1983) and Benston (1985) have questioned whether accounting profit rate data can tell us anything at all about economic performance or its determinants. As Fisher and McGowan put it, "there is no way in which one can look at accounting rates of return and infer anything about relative economic profitability or, a fortiori, about the presence or absence of monopoly profits" (1983, p. 90). If Fisher and McGowan are correct, then the reader need go no further. We think they overstate their case, however.

Fisher, McGowan, and Benston give several reasons for why accounting profits and accounting assets are incorrect measures of their ideal economic counterparts. Fisher and McGowan present several examples suggesting extremely high measurement errors. The basic point they make

is surely valid. Accounting returns *may* deviate from true economic returns by large magnitudes. Whether they do in fact deviate by wide margins is a separate question, however, a question that cannot be answered by presenting hypothetical examples.

The Fisher–McGowan critique focuses upon problems arising in the calculation of the internal rate of return on a single investment project due to variations in the patterns of returns over time and depreciation schedules. As Edwards, Kay, and Mayer (1987, hereafter EKM) point out, however, for most questions in industrial organization, it is the return on assets for the *entire firm* over a particular *interval of time* that is of interest (pp. 32–6).

They derive the following equation that estimates the deviation of the one period accounting rate of return, π_t, from the true economic return of the entire firm over its lifetime, r, as

$$\pi_t = r + (g_a - g_m) + (\pi_t - g_a)\frac{M_{t-1} - A_{t-1}}{A_{t-1}}, \tag{A.1}$$

where A_{t-1} and M_{t-1} are the book value of the firm's assets and their value derived as the present value of future profits and g_a and g_m are the growth rates of these two variables (derived from EKM, 1987, eqs. 2.4, 2.6, and 3.5). We shall use annual observations on π_t to estimate long run projected or permanent profit rates for each firm, π_p, and speeds of adjustment $(1-\lambda)$ to these π_p.

With respect to the two bias terms in (A.1), it is difficult to believe that g_a could permanently exceed or fall short of g_m. Were such the case, the book value of a firm's assets would increasingly diverge from its present value based on future profits. Considering the market value of a firm as an estimate of the latter, one would expect to see the ratio of market to book value converging on zero or infinity, something one does not regularly observe. It is likely, however, that $g_a \neq g_m$ each year but rather sometimes exceeds and sometimes falls short of g_m. Thus, $g_a - g_m$ can be expected to impart a nonsystematic error in the measurement of r using π_t.

If the book value of assets correctly measured the present value of these assets $(A_{t-1} = M_{t-1})$, the last term in (A.1) would be zero. But, as EKM note (p. 35), conservative accounting conventions should result in the book value understating the present value so that $M_{t-1} - A_{t-1}$ is generally positive. The bias introduced by the third term in (A.1) should have the same sign as $\pi_t - g_a$.

We measure π_t for each firm as a deviation of a firm's profit rate in year t from the sample mean for that year. It is likely that $\pi_t - g_a < 0$ for firms with low π_t and $\pi_t - g_a > 0$ for firms with high π_t. An upward bias in π_t as a measure of r is introduced for high π_t's, and a downward bias for low π_t's.

One of the findings of this book will be that some firms have persistently high accounting profit rates and others persistently low accounting rates. Given the likely signs of the third terms in (A.1), it is possible that this finding is at least in part due to systematic bias in the measurement of r using π_t. In some of the country studies, we shall relate π_p estimates to economic variables that are hypothesized to explain r. The support we find for these hypothesized relationships suggests that all of the deviations in long run projected profit rates that we observe are not due to accounting induced biases.

We also measure the speeds of adjustment of π_t to our estimates of permanent profits π_p. We shall interpret these speeds of adjustment as measures of the rapidity with which market forces eliminate short run deviations from a firm's permanent profit levels. Equation (A.1) suggests that short run biases in the measurement of r may also affect these speeds of adjustment. If $g_a - g_m > 0$ for several periods but converged back on zero, then π_t would tend to fall toward r, with the speed of adjustment a measure of the speed with which this component of the accounting bias is eliminated. Although we shall interpret our measures of π_p and $1-\lambda$ as estimates pertaining to underlying economic profits, these possible accounting biases must be kept in mind.

One way to judge the seriousness of the biases from accounting data is to estimate the same types of relationships using data free from or containing smaller biases. Many writers have argued that the market value of a firm is an unbiased measure of a firm's economic performance and have employed Tobin's q as a substitute for accounting rates of return in structure–performance studies. It is defined as the ratio of the market value of the firm to the replacement value of the firm's assets. In comparison to accounting profits over total assets as a measure of economic performance, Tobin's q differs in both the numerator and the denominator. Of these two differences, the change in the numerator is the greatest innovation.

By definition, the market value of the firm is the sum of the market values of its outstanding shares of common stock, debt, and other forms of securities. Each of these in turn should equal the present discounted value of expected future dividends, interest payments, and so on. The argument that the market value of the firm is superior to accounting profits rests on two assumptions, neither of which is obviously true: (1) dividends and interest are paid out of economic profits not accounting profits and (2) the securities markets can accurately predict future economic profits.

The pioneering study in this literature was by Lindenberg and Ross (1981). After an elaborate development of the methodology for measuring