Why Economics must abandon its theory of the firm

Steve Keen, University of Western Sydney,
John Legge, La Trobe School of Management
Geoffrey Fishburn,University of New South Wales
Michael Kelly, Illinois Institute of Technology, &
Mike Honeychurch, University of Melbourne

Abstract: The Marshallian theory of the firm contains previously unrecognised mathematical contradictions which render the theory meaningless. Perfect competition and the welfare ideal of price equal to marginal cost are incompatible with profit maximising behaviour, the comparison of a competitive industry to a monopoly is invalid, and the static profit maximisation principle is false in a dynamic setting. Economics must abandon the theoretical analysis of production based on diminishing marginal productivity, and instead build an empirical microeconomics based on the substantial literature on the behaviour of actual firms.

Keywords: Competition, monopoly, perfect competition, profit maximisation

JEL Classifications: A20, D20, D41, D42, D46, D50, D60, L11, L21

Why Economics must abandon its theory of the firm

Abstract: The Marshallian theory of the firm contains previously unrecognised mathematical contradictions which render the theory meaningless. Perfect competition and the welfare ideal of price equal to marginal cost are incompatible with profit maximising behaviour, the comparison of a competitive industry to a monopoly is invalid, and the static profit maximisation principle is false in a dynamic setting. Economics must abandon the theoretical analysis of production based on diminishing marginal productivity, and instead build an empirical microeconomics based on the substantial literature on the behaviour of actual firms.

Keywords: Competition, monopoly, perfect competition, profit maximisation

JEL Classifications: A20, D20, D41, D42, D46, D50, D60, L11, L21

Why Economics must abandon its theory of the firm

Economic theory is a blend of many intellectual traditions. Walras' general equilibrium clearly provides the core of the discipline, but many other strands of thought contribute to the modern melange: Marshallian microeconomics, von Neumann's game theory, von Neumann-Morgenstern expected utility theory, and so on.

Marshallian microeconomics is the subject of this paper. In the early 20th century, this was the heart of both the pedagogy and research agenda of the discipline. Clearly, the latter is no longer the case, but the former remains true at undergraduate level. Its influence is strong amongst policy economists: those engaged in competition policy generally think in Marshallian terms when considering industry structures. It is also arguable that when those working in research in other strands of economics think of a firm, they think in Marshallian terms.

We argue below that this state of affairs cannot be allowed to continue if economics is to have any pretensions to being a science, whether theoretical, empirical, or both. This is because Marshall's models— specifically the model of perfect competition and the model of monopoly—contain hitherto unrealised mathematical flaws which, when corrected, make the theory devoid of content. So long as the theory of the firm continues to rest on Marshallian foundations and continues to argue that, ideally, price should equal marginal cost, then economics has no theory of the firm.

In this paper, we show using very simple mathematics that:

- Price equals marginal cost is not a profit maximizing equilibrium for a competitive industry;
- Price cannot equal marginal revenue for the individual competitive firm unless either the market demand curve is horizontal, or the output of the firm is zero;
- The sum of the marginal revenue curves of all the competitive firms equals the marginal revenue curve for the market, so that the perfectly competitive price and output levels are identical to the monopoly levels;
- Empirically, a simulated industry converges to the output level at which price equals marginal revenue, regardless of the number of firms in that industry;
- The comparison of monopoly and perfect competition (leaving aside the issue of the validity of the concept of perfect competition itself) is only possible under conditions of identical constant marginal cost;
- While static profit is maximised when marginal revenue equals marginal cost, the rate of growth of profit is not maximised by equating marginal cost to marginal revenue, and therefore the static profit maximising condition does not apply in a dynamic setting.

Since these conclusions are so much at odds with accepted wisdom in economics, it has been put to the authors in the past that their analysis must involve mathematical errors. In order to illustrate that this is not the case, this paper spells its derivations out completely where the mathematics is straightforward, or makes use of the symbolic logic and numerical routines of *Mathematica* and *Mathcad* where the number of steps might lead to confusion.

We conclude with the proposition that the findings of Means et al, hitherto largely ignored by the profession, should become the foundation of microeconomic analysis. Diminishing marginal productivity, the core concept of Marshall’s theory of the firm, will play little or no role in this new microeconomics.

1 Price equals marginal cost is not a profit maximizing equilibrium under perfect competition

The standard approach

The proposition that a competition in a perfectly competitive market results in an output level at which price equals marginal cost is normally established using the following logic. Firstly, total revenue, total cost and profit are defined as functions of quantity:

```
In[1]:= TR@x_D := P@xD x;
        Profit@x_D := TR@xD - TC@xD;
```

(1)

Secondly, given the conditions that the first differential of price is negative and the first differential of total cost is negative, profit is maximised where the first differential of profit is zero:

```
In[3]:= D@Profit@xD, xD Š 0

Out[3]= P@xD + x P¢@xD - TC¢@xD == 0
```

(3)

Thirdly, the assumption is made that because the individual competitive firm is so small relative to the entire industry, the differential of price with respect to the output of a single firm is zero (this assumption is often justified by reformulating the second term in equation 3 in terms of the price elasticity of demand for a single firm; we consider this issue further below):

```
In[4]:= P '@x_D := 0;
```

(4)

Therefore, for the individual competitive firm, profit is maximised when marginal cost equals price:

```
In[5]:= D@Profit@xD, xD Š 0

Out[5]= P@xD - TC¢@xD == 0
```

(5)

We argue that this logic is flawed from the moment the assumption is made that $P'[x] = 0$. If this assumption is true, then it should be provable in its own right; if it is false, then it should be an approximation that has only a minimal impact on the analysis. We demonstrate below that the assumption is false, and that its impact on the analysis is substantial. Without it, the theory of perfect competition collapses, and the outcomes for a competitive market are provably identical to those for a monopoly.

Without assumptions (Stigler 1957)

Firstly, we remove the definition of price to eliminate assumptions about its differential (and reintroduce the definition of total revenue for technical reasons in *Mathematica*):

```
In[6]:= Remove@PD;
        TR@x_D := P@xD x;
```

(6)

Next, the alleged perfectly competitive and monopoly output levels are defined: *qpc* as the level at which price equals marginal cost and *qm* as the level at which marginal revenue equals marginal cost:

```
In[8]:= qpc •: TC'@qpcD := P@qpcD;
        qm •: TC'@qmD := TR'@qmD;
```
(7)

The first differential of profit is as expected:

```
In[10]:= Profit'@qD

Out[10]= P@qD + q P¢@qD - TC¢@qD
```
(8)

Now we consider the assertions of neoclassical theory. Firstly we input the condition that the profit for a monopoly is maximised at *qm*. The answer from the symbolic engine is oracular:

```
In[11]:= D@Profit@qmD, qmD Š 0

Out[11]= True
```
(9)

When given the customary perfect competition condition for profit maximisation for a single firm, *Mathematica* no longer answers “True” but reduces to its essence, that the second term in (8) must be zero:

```
In[12]:= D@Profit@qpcD, qpcD Š 0

Out[12]= qpc P¢@qpcD == 0
```
(10)

Since *qpc* is obviously positive, the perfect competition result that price equals marginal cost is possible if and only if $P'[qpc]=0$. This may be taken as reassuring by most economists, since the proposition that the slope of the demand curve facing the individual competitive firm is zero is a core assumption of the theory of perfect competition. However, Stigler demonstrated over 40 years ago that this assumption is mathematically incorrect (Stigler 1957: 8). His expression of this was succinct:

$$\frac{dP}{dq_i}=\frac{dP}{dQ}\frac{dQ}{dq_i}=\frac{dP}{dQ} \tag{11}$$

Stigler was implicitly applying a fundamental assumption of the model of perfect competition, that firms do not react to the behaviour of other firms. Stating this explicitly, we define *q* as the output of a single firm and Q_R as the output of all other firms in the industry. The assumption of atomism then gives us that $\frac{dQ_R}{dq_i}=0$. Hence:

$$\begin{aligned}
\frac{dP}{dq_i} &= \frac{dP}{dQ}\frac{dQ}{dq_i}\\
&= \frac{dP}{dQ}\frac{d}{dq_i}(q_i+Q_R)\\
&= \frac{dP}{dQ}\left(\frac{d}{dq_i}q_i+\frac{d}{dq_i}Q_R\right)\\
&= \frac{dP}{dQ}(1+0)\\
&= \frac{dP}{dQ}<0
\end{aligned} \tag{12}$$

The assumption that $\frac{dP}{dq_i}=0$ is therefore incompatible with the assumption that the market demand curve is negative. Given $\frac{dP}{dQ}<0$, then the demand curve facing an individual firm is also downward sloping. The demand curve for the individual firm can only be horizontal if the market demand curve itself is horizontal.

The elasticity defence

Stigler derived the proposition that $\frac{dP}{dq_i}=\frac{dP}{dQ}$ in the course of developing a reformulation of the relationship of marginal revenue for the individual firm to price as

$$MarginalRevenue = Price + \frac{Price}{NumberofSellers \times MarketElasticity} \tag{13}$$

where he argued that "this last term goes to zero as the number of sellers increases indefinitely" (Stigler 1957: 8). If this were true, then for the very large number of atomistic firms that are necessary for the model of perfect competition, "marginal revenue equals price" would be strictly false, but a reasonable approximation to the truth. However, Stigler's superficially convincing argument contains a serious if understandable error. He began with the marginal revenue function for a single firm, using the equality he had previously established between $\frac{dP}{dq_i}$ and $\frac{dP}{dQ}$:

$$\frac{d}{dq_i}(P\times q_i)=P+q_i\frac{dP}{dQ} \tag{14}$$

and then worked with the simplification of identical output levels $(Q=nq)$ to perform the substitutions

$$\begin{aligned}\frac{d}{dq}(P\times q)&=P+q\frac{dP}{dQ}\\&=P\left(1+\frac{Q}{n}\frac{1}{P}\frac{dP}{dQ}\right)\\&=P+\frac{P}{n\times E}\end{aligned} \tag{15}$$

where $E=\frac{P}{Q}\frac{dQ}{dP}$ is the market elasticity of demand.

While it might seem reasonable to presume that E and P are constants, so that $\frac{P}{n\times E}\to 0$ as $n\to\infty$, it should be obvious from the first line of (15) that this is true if and only if $q\to 0$ as $n\to\infty$ or $\frac{dP}{dQ}=0$. If instead q remains constant and $\frac{dP}{dQ}<0$, then $\frac{P}{n\times E}$ will not go to zero as $n\to\infty$, because E will fall in step with the rise in n. This can be made obvious by treating firm size $q=\frac{Q}{n}$ as a constant and substituting this into Stigler's final line:

$$\begin{aligned}P+\frac{P}{n\times E}&=P+\frac{P}{\frac{Q}{q}\times\frac{P}{Q}\frac{dQ}{dP}}\\&=P+\frac{P}{\frac{P}{q}\frac{dQ}{dP}}\end{aligned} \tag{16}$$

For $\frac{P}{n\times E}=\frac{P}{\frac{P}{q}\frac{dQ}{dP}}$ to go to zero as $n\to\infty$, either $\frac{P}{q}\to\infty$ (which is only possible if $q\to 0$) or $\frac{dQ}{dP}$ must equal infinity. The first condition is clearly nonsense, while $\frac{dQ}{dP}=\infty$ repeats the proposition shown above, that price can equal marginal revenue only if the market demand curve is horizontal.

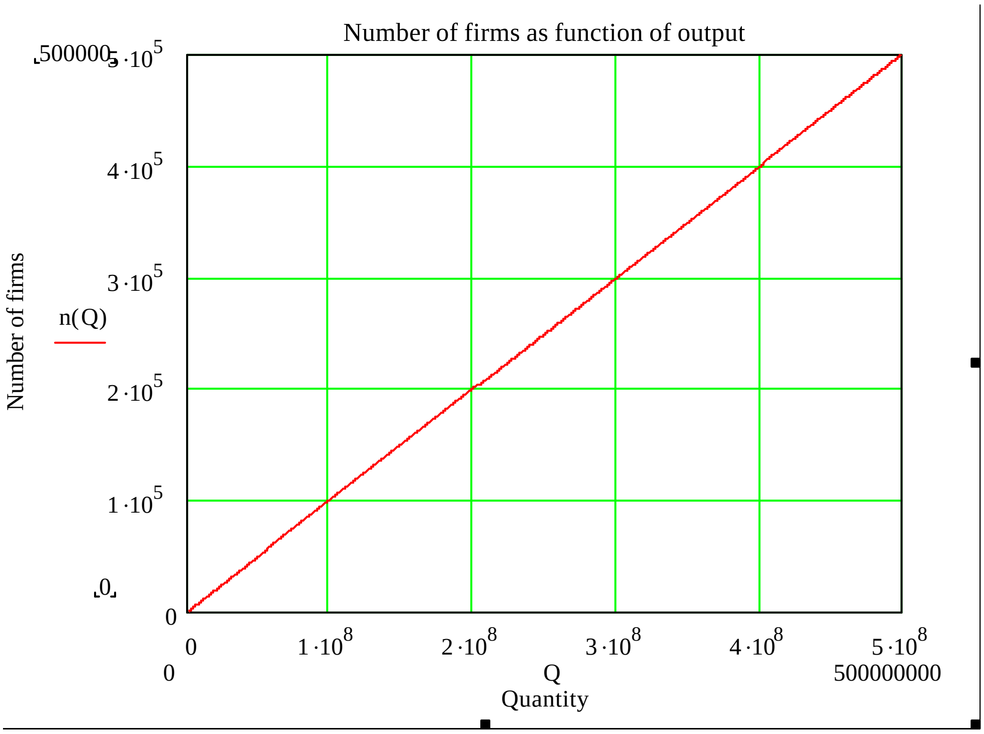
In fact, given a constant firm size q, $\frac{P}{n \times E}$ is a constant: it is completely independent of the term n:

$$\begin{aligned}\frac{P}{n\times E} &= \frac{P}{n\times \frac{P}{Q}\frac{dQ}{dP}} \\ &= \frac{P}{n\times \frac{P}{q\times n}\frac{dQ}{dP}} \\ &= \frac{1}{\frac{1}{q}\frac{dQ}{dP}} \\ &= q\frac{dP}{dQ}\end{aligned} \qquad (17)$$

This, of course, was the first line in Stigler's derivation. The remainder of his workings after this were irrelevant, but unfortunately served to shield economists from a serious logical flaw in economic theory. To emphasise the irrelevance of Stigler's reworking of the term for the marginal revenue function for a single firm, we give a numerical example below in which $\frac{P}{n \times E}$ is calculated for a given demand curve and firm size as $n \to \infty$.

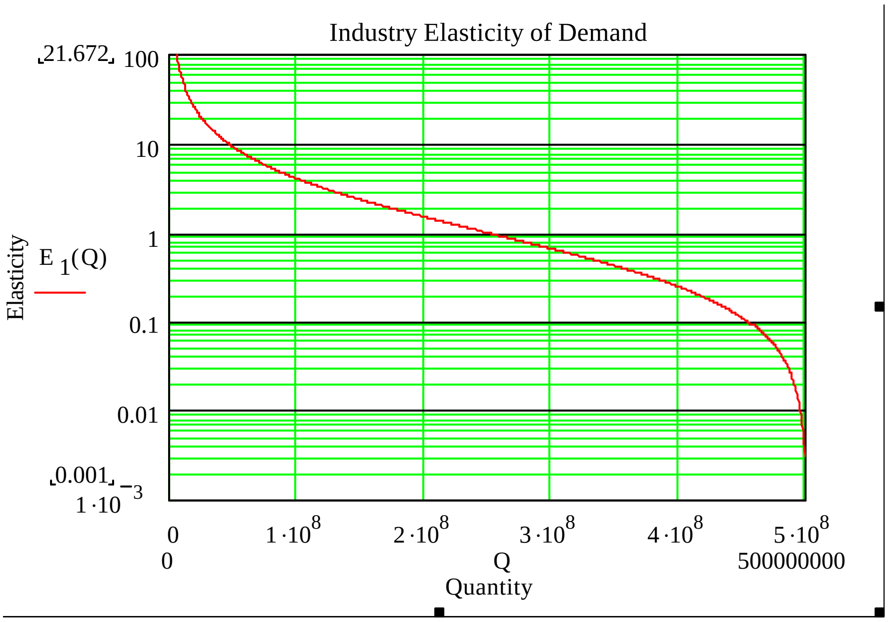
Consider an industry with a linear demand curve given by $P = 100000 - \frac{1}{5000}Q$ and identical firms each with an equilibrium output of $q = 1000$ units. Then n is a linear function of Q (Figure 1):

Figure 1. With constant demand and constant firm size, n rises as Q rises



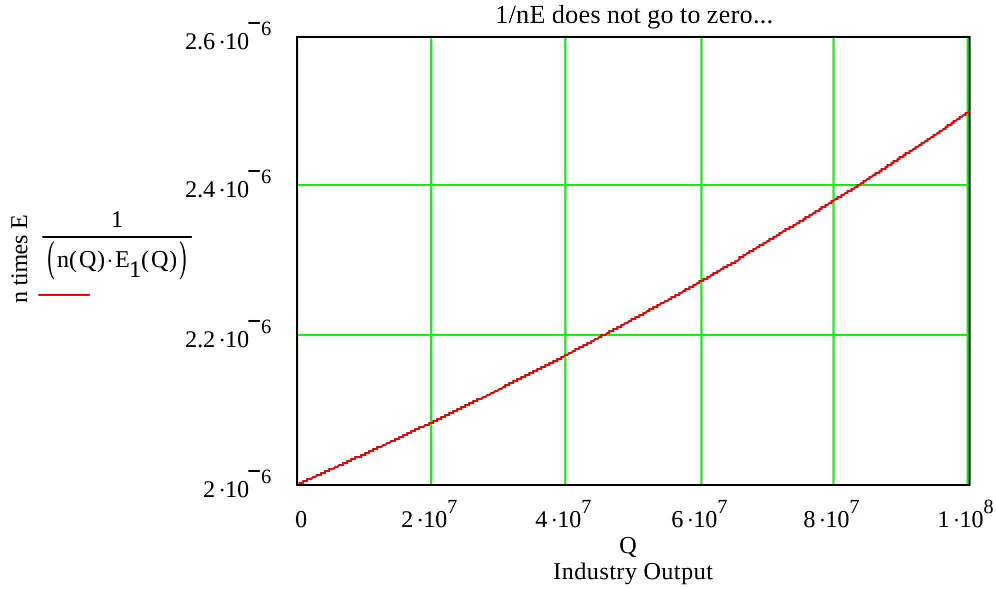
E, on the other hand, is a nonlinear decreasing function of Q (Figure 2):

Figure 2. E falls nonlinearly as Q rises



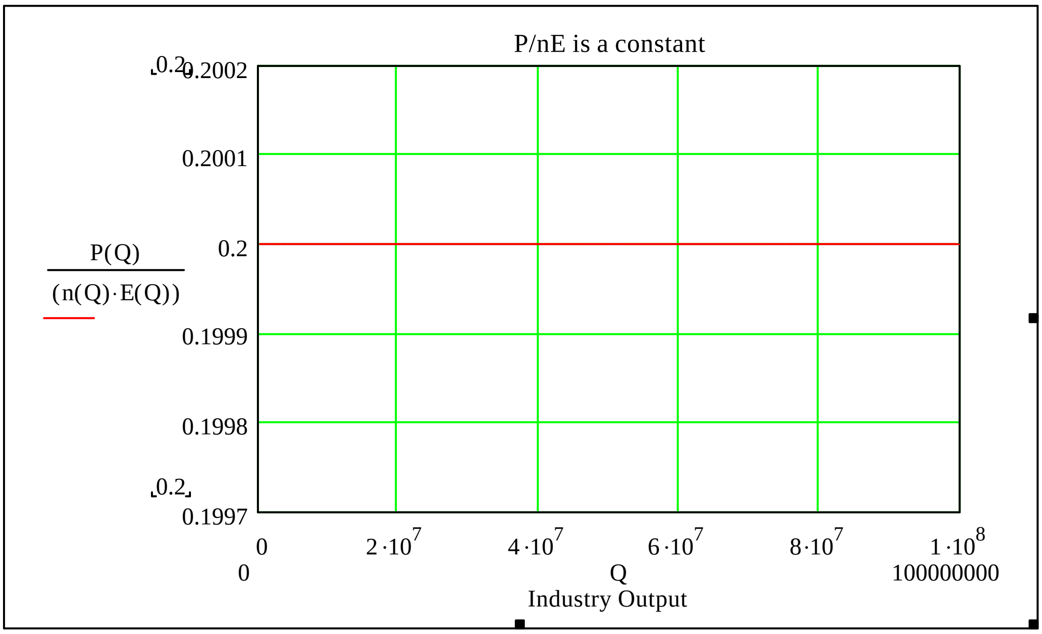
Far from going to zero as n increases, the inverse product $\frac{1}{n \times E}$ is an increasing function of Q, and therefore of n (Figure 3):

Figure 3. The term $\frac{1}{n \times E}$ rises as Q (and n) rises



As is already obvious from the algebra, $\frac{P}{n \times E}$ is a constant (Figure 4):

Figure 4. $\frac{P}{n \times E}$ **is a constant**



The number of firms in an industry per se thus has no effect on the term $\frac{P}{n \times E}$, and therefore marginal revenue for the individual firm is always less than price, regardless of the number of firms in the industry.

It remains true that the marginal revenue for the *i*th firm with respect to a change in its own output is very close to the market price and much greater than market marginal revenue:

$$P > P + \frac{P}{n \times E} \gg P + \frac{P}{E} \tag{18}$$

But the gap between market price and own-output marginal revenue is independent of the number of firms in the industry. Taking the general example of a linear demand curve, with n identical firms each producing q units of output, we have:

$$\begin{aligned} P &= a - bQ \\ Q &= n \times q \\ E &= \frac{P}{Q}\frac{dQ}{dP} = \frac{a-bQ}{Q} \times \frac{-1}{b} \\ P - MR_i &= \frac{P}{n \times E} = \frac{a-bQ}{n \times \frac{a-bQ}{Q} \times b} \\ &= \frac{a-bQ}{\frac{Q}{q} \times \frac{a-bQ}{Q} \times b} = qb \end{aligned} \tag{19}$$

The difference between market price and own-output marginal revenue is thus a constant, regardless of the number of firms in the industry, and is equal to the output level of the representative firm times b—the linear equivalent to Stigler's more general first line in (15).

Since own-output marginal revenue is always less than market price for the competitive firm, it should be evident that the condition that "price equals marginal cost" cannot be a profit maximising equilibrium. Any firm producing at an output level where its marginal cost equals the market price will be making a loss on the last units produced, because the marginal cost of producing them will exceed the marginal revenue gained from selling them. The firm will thus increase its profit if it reduces its output.

The question remains, if the industry begins at the output level where price equals marginal cost, and firms therefore have an incentive to reduce output, what will this reduction in output sum to over all n firms in the industry? In this process we have to take account of the feedback effects between different firms, since the reduction in output by one firm (which increases its profits, reduces its output and increases market price) will impact on the profitability of all other firms (who benefit from the increase in market price while not reducing their output).

Disaggregated profit maximising

We consider (without loss of generality) a market facing a linear demand curve, with n firms. Initially, we will allow firms to be of varying sizes, so that the output of the ith firm is q_i and its cost function is $TC_i(q_i)$.

Price is a linear function of market output, and market output is the sum of the output of the n firms in the industry. Profit for the jth firm is therefore:

$$Profit_j = P(Q)q_j - TC_j(q_j) = P\left(\sum_{i=1}^{n} q_i\right)q_j - TC_j(q_j) = (a - bQ)q_j - TC_j(q_j) \quad (20)$$

The differential of profit with respect to output q_j is:

$$\frac{d}{dq_j} Profit_j = \frac{d}{dq_j}\left(P(Q)q_j\right) - MC_j(q_j) \quad (21)$$

Expanding the differential of total revenue:

$$\begin{aligned}
\frac{d}{dq_j}\left(P(Q)q_j\right) &= \frac{d}{dq_j}\left(q_j \times (a - bQ)\right) \\
&= \frac{d}{dq_j}\left(q_j \times \left(a - b\sum_{i=1}^{n} q_i\right)\right) \\
&= \frac{d}{dq_j}\left(aq_j - b \times q_j \times \sum_{i=1}^{n} q_i\right) \\
&= a - b \times \frac{d}{dq_j}\left(q_j \times \left(q_1 + q_2 + \ldots + q_j + \ldots + q_n\right)\right) \\
&= a - b \times \frac{d}{dq_j}\left(q_1 q_j + q_2 q_j + \ldots + q_j^2 + \ldots + q_n q_j\right) \\
&= a - b \times \left(q_1 + q_2 + \ldots + 2q_j + \ldots + q_n\right) \\
&= a - b \times \sum_{i=1}^{n} q_i - bq_j \\
&= a - bQ - bq_j = P(Q) - bq_j
\end{aligned} \quad (22)$$

This confirms the result derived from Stigler's formula for the own-output elasticity of demand.

It might be felt that for large n the quantity bq_j is sufficiently small that it can be ignored, and that therefore approximation can be used to justify the standard argument that marginal revenue equals price for the individual competitive firm. The number bq_j will indeed be small for any industry which approaches the perfect competition condition of a large number of firms that do not directly react to each other's behaviour, since though each q_j could be quite large (say 40,000 bushels of wheat), the coefficient b would be extremely small. However, what may appear to be a valid approximation for the individual firm does not remain so when one aggregates from the single firm to the entire industry. It can easily be shown that the sum of the marginal revenue curves facing each individual firm is equivalent to the marginal revenue curve facing the entire industry—and thus that the competitive industry will produce the same output level as a monopoly. Since monopoly output (where marginal revenue equals marginal cost) and the alleged perfectly competitive output level (where price equals marginal cost) are substantially different, it is therefore not valid to ignore the term bq_j above.

Marginal revenue for a monopolist facing a linear demand curve is:

$$MR = \tfrac{d}{dQ}(P \times Q) = \tfrac{d}{dQ}(aQ - bQ^2) = a - 2bQ \tag{23}$$

The sum of the marginal revenue functions for the n competitive firms is not the same as marginal revenue for the industry, because of the feedback effects mentioned above. Instead, sum of the marginal revenue functions can be shown to be equal to the industry marginal revenue plus *n-1* times market price:

$$\begin{aligned}
\sum_{i=1}^{n}\left(\frac{d}{dq_i}(P \times q_i)\right) &= \sum_{i=1}^{n}\left(P + q_i \frac{d}{dq_i}P\right) \\
&= nP + \sum_{i=1}^{n}\left(q_i \frac{d}{dQ}P\right) \\
&= nP + \frac{d}{dQ}P \sum_{i=1}^{n} q_i \\
&= nP + Q\frac{d}{dQ}P \\
&= (n-1)P + \left(P + Q\frac{d}{dQ}P\right) \\
&= (n-1)P + MR
\end{aligned} \tag{24}$$

We now use this relation and expand $\sum_{i=1}^{n}(a - bQ - bq_i)$:

$$\begin{aligned}
(n-1)P + MR &= \sum_{i=1}^{n}(a - bQ - bq_i) \\
&= na - nbQ - b\sum_{i=1}^{n} q_i
\end{aligned} \tag{25}$$

$$(n-1)P + MR = nP - bQ$$

$$MR = P - bQ = a - 2bQ$$

Thus the marginal revenue curve for a perfectly competitive industry is identical to that for a monopoly. Since we still accept the condition that profit is maximised by equating marginal revenue and marginal cost, it follows that a competitive industry with an aggregate marginal cost curve equivalent to the marginal cost curve for a monopolist will produce the same output as the monopolist, and sell it for the same price. There is thus no formal difference between the price, quantity and welfare outcomes of the models of competition and monopoly.

An empirical example illustrates this, and also indicates another problem with the theory of the firm: the output and price decisions of a monopoly cannot be meaningfully compared to those of a competitive market under conditions of diminishing marginal productivity.

An empirical example

We take our lead from Friedman 1953, in which he argued for accepted economic theory on the basis that while economic agents do not necessarily consciously solve the optimising formulae taught to students, any agent that did not behave "as if" it was doing this would fail. Giving the example of expert billiard players, Friedman argued that:

> "Excellent predictions would be yielded by the hypothesis that the billiard player made his shots *as if* he knew the complicated mathematical formulas …, could make lightning calculations from the formulas, and could then make the balls travel in the direction indicated by the formulas. Our confidence in this hypothesis is not based on the belief that billiard players, even expert ones, can or do go through the process described; it derives rather from the belief that, unless in some way or other they were capable of reaching essentially the same result, they would not in fact be *expert* billiard players." (Friedman 1953: 21)

We now pose the question: if a firm simply "groped" for a level of output on the basis of whether its profit rose or fell, would its output level converge to where price equals marginal cost? Our hypothetical industry has the following linear demand curve:

$$P(Q) = 110 - \frac{1}{100000000}Q \qquad (26)$$

For reasons that are explained below, our first example uses a constant cost production function, with of course the same function for both competitive firms and a monopoly (our second example uses a rising marginal cost production function):

$$TC(q) = 50q \qquad (27)$$

With these parameters, the profit maximising level of output for a monopoly is *3,000,000,000* units at a price of *80*, while the alleged level for a competitive industry is *6,000,000,000* at a price of 50.

Our program (see Figure 5) takes as its argument the number of firms in an industry, and then assigns a random initial production level to each firm. Each firm then works out its profit, and makes a random change in its output level. If this change in output increases its profit, the firm continues to change its output in the same direction in the next iteration, but by a smaller amount (governed by the function *Sig*, which generates a logistic curve from the arguments of a maximum and minimum value, an x-coordinate at which it lies half way between those values, and a slope at that point; see

Equation 28). The same process is repeated on subsequent iterations (to a maximum of 50 iterations). No firm takes any notice of the behaviour of any other firm. If Marshallian theory is correct, and if the concept "many firms" can be given any empirical substance, then the point of convergence of output and price should be in some sense a function of the number of firms, with a single firm converging to the monopoly level and a many firm industry converging towards the perfect competition level.

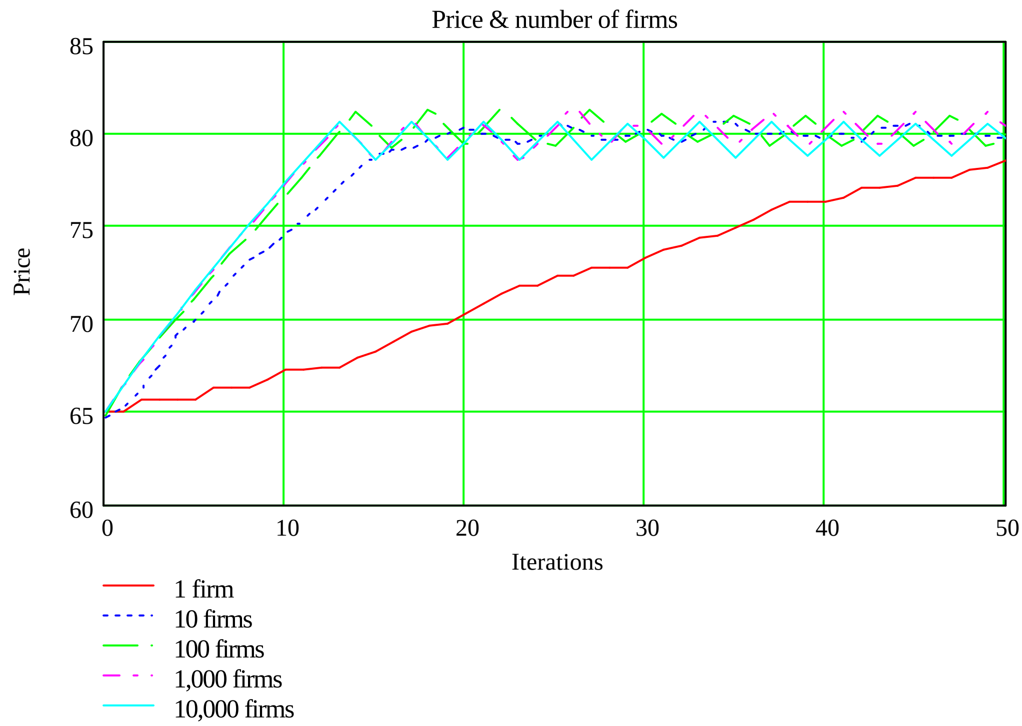
Figure 5. Profit-maximising by random movements in output

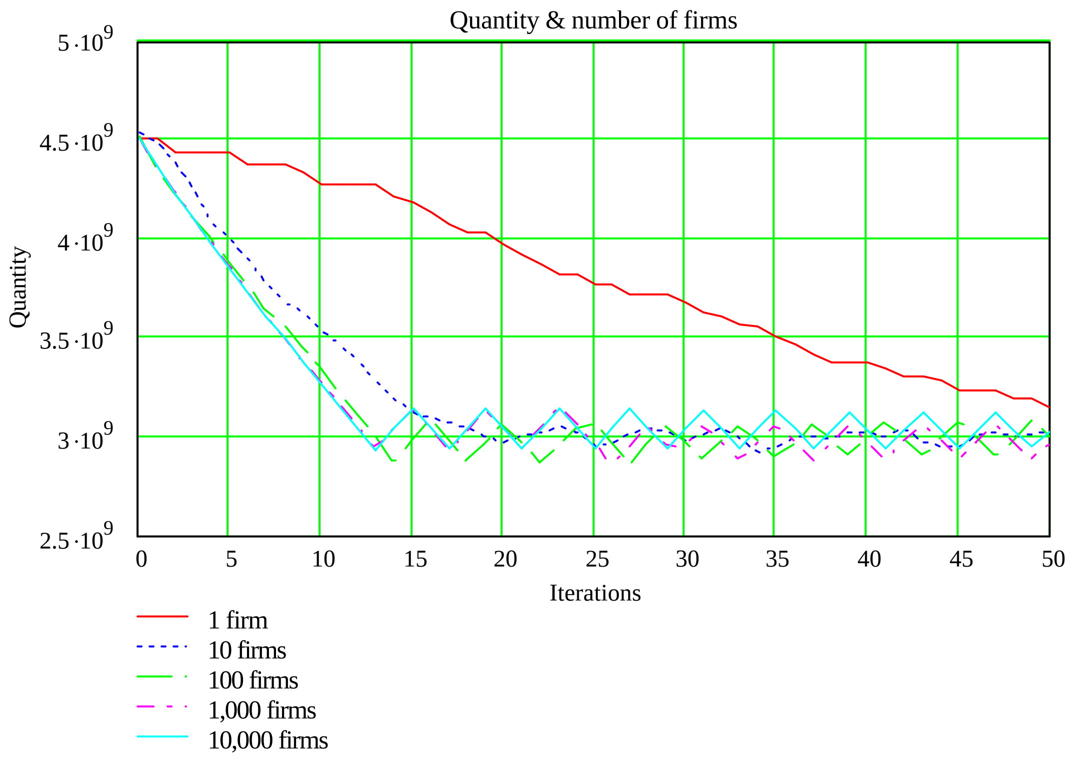
$$
\begin{aligned}
\mathrm{Mkt}(n) := \quad & F \leftarrow \mathrm{round}\left(\mathrm{rnorm}\left(n, \frac{Q_M + q_{pc}}{2\cdot n}, \frac{\sqrt{q_{pc}}}{n}\right)\right) \\
& \mathrm{Profit}_{ini} \leftarrow P\left(\sum F\right)\cdot F - tc(F) \\
& dq_{ran} \leftarrow \mathrm{round}\left[\left[\mathrm{runif}\left[n, -\left[\frac{Q_M}{(n+1)^2}\right], \frac{Q_M}{(n+1)^2}\right]\right]\right] \\
& F \leftarrow F + dq_{ran} \\
& \mathrm{Profit}_{ran} \leftarrow P\left(\sum F\right)\cdot F - tc(F) \\
& dq \leftarrow \overrightarrow{\left(\mathrm{sign}\left(\mathrm{Profit}_{ran} - \mathrm{Profit}_{ini}\right)\cdot dq_{ran}\right)} \\
& \text{for } i \in 0..50 \\
& \quad \mathrm{Profit}_{pre} \leftarrow \overrightarrow{\left(P\left(\sum F\right)\cdot F - tc(F)\right)} \\
& \quad F \leftarrow F + dq \\
& \quad \mathrm{Profit}_{post} \leftarrow \overrightarrow{\left(P\left(\sum F\right)\cdot F - tc(F)\right)} \\
& \quad dq \leftarrow \left[\overrightarrow{\mathrm{sign}\left[\left(\mathrm{Profit}_{post} - \mathrm{Profit}_{pre}\right)\cdot dq\right]}\right]\cdot\left[\mathrm{Sig}\left[0, \frac{1}{10+\sqrt{i}}, \frac{1}{2\cdot\left(10+\sqrt{i}\right)}, \frac{1}{2+\sqrt{i}}, \mathrm{runif}(n,-2,2)\right]\cdot\frac{Q_M}{(n+1)^2}\right] \\
& \quad \mathrm{Price}_i \leftarrow P\left(\sum F\right) \\
& \quad \mathrm{Quantity}_i \leftarrow \sum F \\
& \mathrm{augment}(\mathrm{Price}, \mathrm{Quantity})
\end{aligned}
$$

$$Sig(Min, Max, H, S, x) = \frac{Max - Min}{1 + e^{4S\frac{H-x}{Max-Min}}} + Min \qquad (28)$$

As Figure 6 shows, price converges to the monopoly level no matter how many firms are in the industry.

Figure 6. Price and output for different numbers of firms



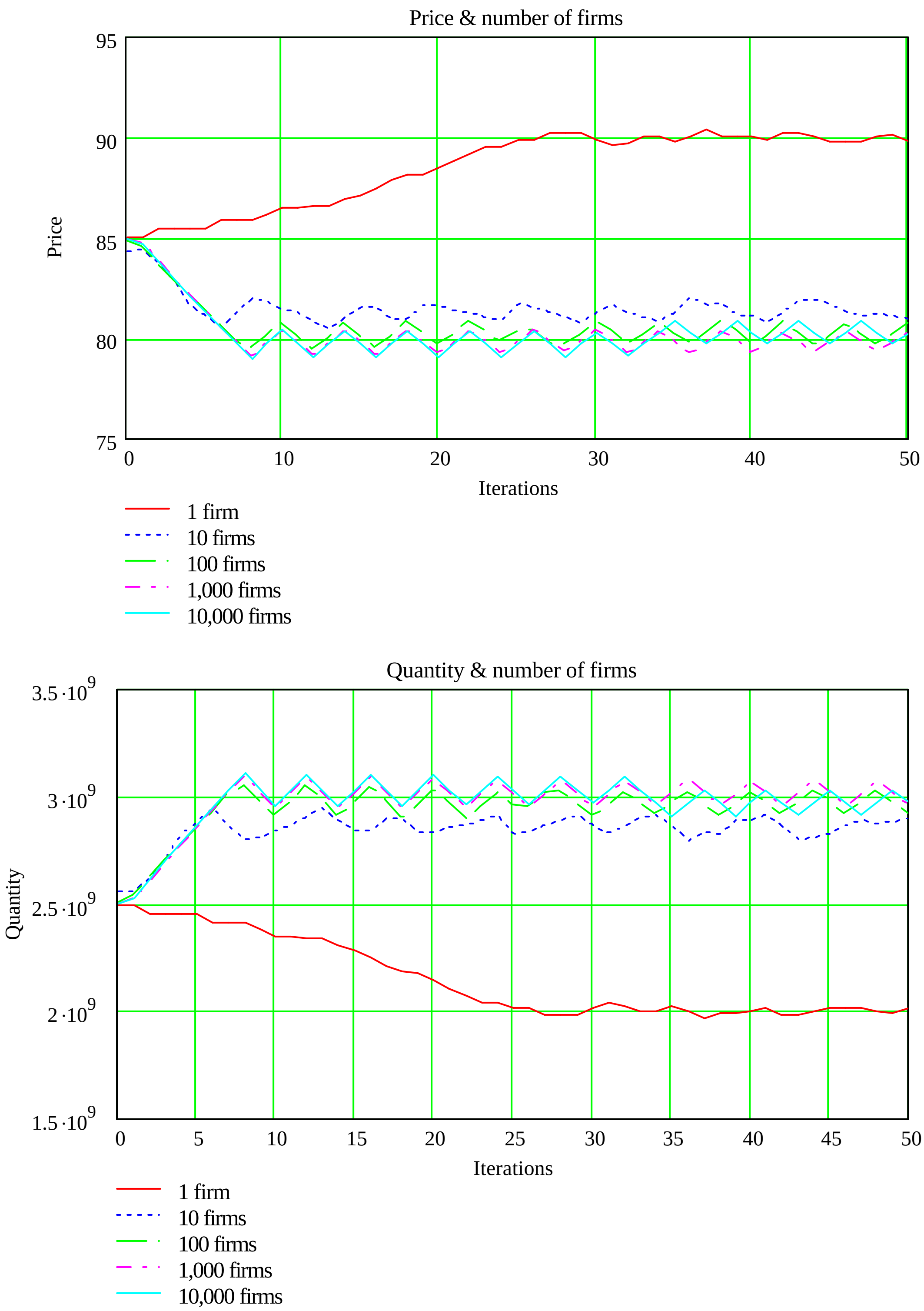


This result may be thought to be an artifact of the assumption of constant costs. Our second empirical example uses a cost function with rising marginal cost:

$$TC(q) = 50q + 5 \times 10^{-9} q^2 \quad (29)$$

With this cost function, the monopoly levels of output and price are respectively *2,000,000,000* and *90*, while the alleged competitive levels are *3,000,000,000* and *80*. At first glance, Figure 7 seems to confirm the conventional theory: the monopoly converges to the monopoly price and quantity, while the price for any larger number of firms apparently converges to the competitive levels.

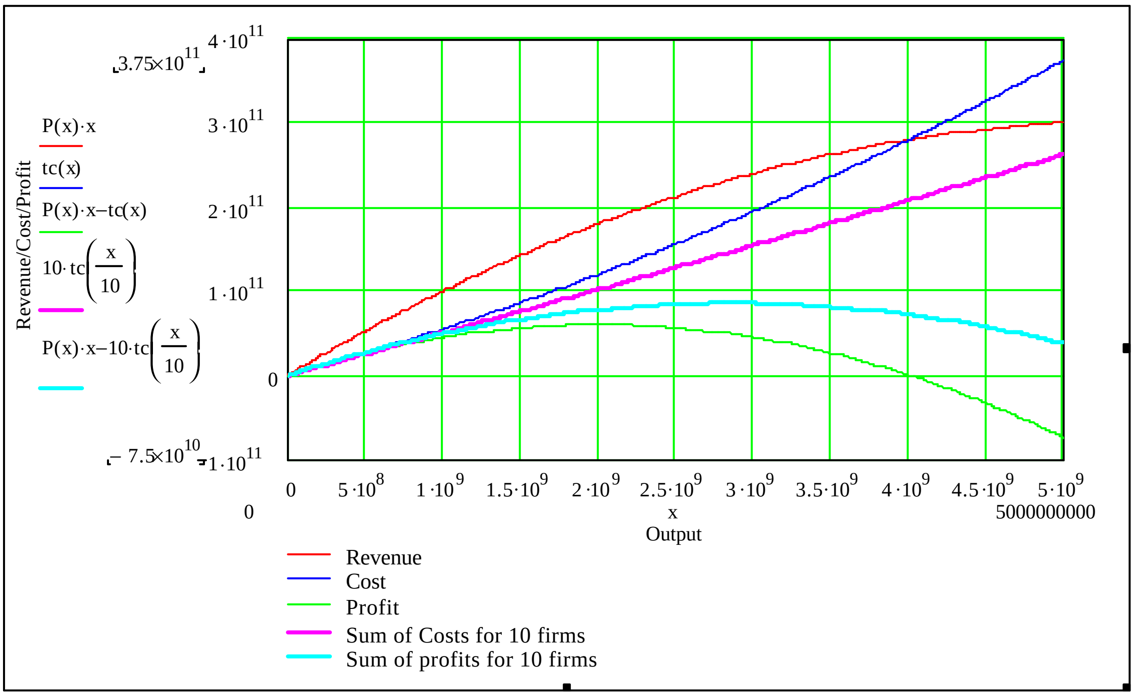
Figure 7. Price and output for different numbers of firms with rising marginal cost



However, this apparent confirmation of the theory in fact points to another flaw in the Marshallian model. The results differ, not because the monopoly produces where marginal cost equals marginal revenue while a competitive industry produces where marginal cost equals price, but because given diminishing marginal productivity, the cost function for the monopoly necessarily differs from the cost curves for an industry with more than one producer.

Figure 8 compares the total cost and profit curves for a single firm industry and a ten firm industry where each firm has the same total cost function. It is apparent that the sum of the total cost curves for the ten firm industry is lower than the total cost curve for the single firm industry. The higher level of output of the multi-firm industry occurs, not because of a difference in profit maximisation principles, but because of lower costs for the many firm instance than for the monopoly.

Figure 8. Total revenue and aggregate total cost functions for different numbers of firms



This problem arises because, in order to make a definitive comparison between monopoly and a competitive market, the marginal cost curve for the monopoly must be identical to the sum of the marginal cost curves for the competitive market. If the cost curves differ, then it is quite possible for a monopoly to produce a larger quantity at a lower price than a competitive industry because of lower costs. It is easily shown that this condition of identical marginal cost functions is only possible where all firms have the same constant marginal cost production function.

Monopoly and multi-firm marginal cost curves must differ under diminishing marginal productivity

The condition that the marginal cost curve of a single firm and the sum of marginal cost curves for two or more firms must be identical is simultaneously the condition that their marginal products must be identical—since differences in marginal product are the only allowable source of differences in per unit cost. If the marginal product curves are identical, then total product curves can only differ by a constant. If we take labour as our variable factor of production, then output is zero at zero input, and the constant can be set to zero.

We can put this condition upon the production functions of the monopoly and a multiple-firm industry into the following form: given the same number of variable inputs, the output of the single firm must be identical to the sum of the outputs of the other firms. If we consider (without loss of generality) a multiple firm industry with n firms each employing x workers, then the output of these n firms must be equal to the output of the single monopoly firm employing nx workers. Using f for the production function of the competitive firms, and g for the production function of the monopoly, this condition is Euler's Equation (with a constant of zero).

$$n \times f(x) = g(n \times x) \tag{30}$$

As is easily shown, the only production function that satisfies this condition is one that yields identical constant marginal product, and therefore identical constant marginal cost. Using *Mathematica*, we differentiate both sides of this identity with respect to n:

```
In[47]:= D@n f@xD Š g@n xD, nD
Out[47]= f@xD == x g£@n xD
```

(31)

We next divide the differential of each side of the identity by the identity:

```
In[48]:= D@n f@xD, nD • HHn f@xDLL Š D@g@n xD, nD • g@n xD
Out[48]= 1/n == x g£@n xD / g@n xD
```

(32)

We then divide both sides by x:

```
In[49]:= D@n f@xD, nD • HHn f@xDLL • x Š HD@g@n xD, nD • g@n xDL • x
Out[49]= 1/(n x) == g£@n xD / g@n xD
```

(33)

We now know that the production function g must satisfy this relation. Making the substitution $u = nx$, we find that g must be linear:

```
In[50]:= DSolve@Hg'L@uD Š g@uD • u, g@uD, uD
Out[50]= 88g@uD ® u C@1D<<
```

(34)

Solving for f, we find that it must be the same linear function:

```
In[51]:= Solve@n f@xD Š C@1D n x, f@xD, fD
Out[51]= 88f@xD ® x C@1D<<
```

(35)

Thus only in the case of constant identical marginal costs will the marginal cost curve of a single firm be identical to the sum of the marginal cost curves of two or more firms. As soon as diminishing marginal productivity is introduced, the comparison of monopoly and any other industry structure can no longer be definitive, since their cost functions must differ. Yet diminishing marginal productivity is the cornerstone of the Marshallian analysis of production. We conclude that one other aspect of the conventional theory—the proof that the welfare outcome under perfect competition is preferable to that under monopoly—is vacuous, even if the assumption that perfectly competitive firms face a horizontal demand curve is not challenged.

Given how firmly the assumption that the demand curve is horizontal for a firm in a competitive industry is held by economists, below we provide one final proof that this is erroneous, and that price equals marginal cost cannot be a profit maximising equilibrium for any market structure.

One final proof: a Taylor series expansion

Consider what happens to the profit of an individual firm that is currently producing at the alleged profit-maximising level—where marginal cost equals marginal revenue for a monopoly, and where marginal cost equals price for a competitive firm—when it alters its output by a small amount δq.

We restrict ourselves to linear marginal cost and market demand curves by setting the second differential of price to zero and the second differential of total cost to a constant. This is done simply to remove any confusion arising from third order and higher terms in the Taylor's series expansions we use below; the generality of our results is not affected by this restriction to linear functions:

```
In[18]:= P''@x_D := 0;
         TC''@x_D := d;
         TR@x_D := P@xD x;
         Profit@x_D := TR@xD - TC@xD;
         qpc •: TC'@qpcD := P@qpcD;
         qm •: TC'@qmD := TR'@qmD;
```

(38)

The Mathematica function “Series” produces a Taylor series expansion of the given function; “Normal” drops all terms above the specified power of expansion—2 in this case—and “Simplify” cancels like terms. For monopoly, the difference between profit at the output level $q_m + \delta q_m$ and profit at q_m is given by:

```
In[25]:= Simplify@Normal@Series@Profit@qm + dqmD, 8dqm, 0, 2<DDD - Profit@qmD

Out[25]= dqm^2 J- d/2 + P¢@qmDN
```

(39)

Since the market demand curve is assumed to be negatively sloped, $P'[qm]$ is negative; since marginal cost is assumed to rise, $-\frac{d}{2}$ is also negative. $\delta q_m^2$is positive whether δq_m is positive or negative. Therefore any change in output by a monopoly from the level at which marginal cost equals marginal revenue will cause a fall in profit. A profit maximising monopoly will therefore produce q_m, the output level at which marginal cost equals marginal revenue. This confirms the theory.

Applying precisely the same operations to perfect competition reaches a quite different result:

```
In[26]:= Simplify@Normal@Series@Profit@qpc + dqpcD, 8dqpc, 0, 2<DDD - Profit@qpcD

Out[26]= qpc dqpc P¢@qpcD + dqpc^2 J- d/2 + P¢@qpcDN
```

(40)

The second order term is identical to that for monopoly, and has an identical effect—that it reduces profit for any change of output from the level q_{pc}. However, there is a first order term that was absent in the monopoly result, and this term dominates the second term. Its first component q_{pc} is positive; the sign of its second term δq_{pc} is negative if output is reduced below the level at which price equals marginal cost; while its third term is categorically negative if market demand is a declining function of quantity.

For $\delta q_{pc} < 0$, this gives us a positive times a negative times a negative: the firm's profit level will rise if it reduces its output. Price equals marginal cost is therefore not a profit maximising equilibrium. The competitive firm will increase its profit by reducing its output below the level at which price equals marginal cost. The output level where price equals marginal cost does not maximise profits, but instead involves selling at a loss the output above the level at which marginal revenue equals marginal cost.

2 Profit maximisation and time

It should be evident that little remains of the conventional theory beyond the simple statement that profit is maximised by equating marginal cost and marginal revenue. However even this simple proposition is open to challenge, when one steps outside Marshall's static model into the real dynamic world. In general, the profit earned by a firm is a function, not simply of quantity, but also of time, geographic reach, and many other variables.

If we simply consider time, we must also account for the fact that the quantity produced is itself a function of time: that is, at any point in time, a firm has the possibility of choosing any output level between zero and its installed capacity at that time. Equally, price and costs are functions of time as well as quantity: demand alters over time as well as with respect to quantity, and total cost alters with respect to time (due to changing production technologies, changing input costs, etc.) as well as with respect to quantity. A complete dynamic specification of profit is therefore:

```
In[29]:= TR@x_@y_D, y_D := P@x@yD, yD x@yD;
         Profit@x_@y_D, y_D := TR@x@yD, yD - TC@x@yD, yD;
```

(41)

Given a dynamic definition of profit, it is reasonable to consider that the firm's objective is not simply to achieve the maximum rate of profit with respect to quantity produced at a single point in time, but must to some degree involve achieving the greatest rate of growth of profit over time. Dropping the detail of functional dependence on $q(t)$ and t for expositional clarity, and substituting in marginal cost and marginal revenue as conventionally defined yields, the total differential of profit is obviously:

$$\begin{aligned} dProfit &= \frac{\partial Profit}{\partial t}dt + \frac{\partial Profit}{\partial q}dq \\ &= \frac{\partial (Pq - TC)}{\partial t}dt + \frac{\partial (Pq - TC)}{\partial q}dq \\ &= \left(q\frac{\partial P}{\partial t} + P\frac{\partial q}{\partial t} - \frac{\partial TC}{\partial t}\right)dt + \left(\left(P\frac{\partial q}{\partial q} + q\frac{\partial P}{\partial q}\right) - \frac{\partial TC}{\partial q}\right)dq \\ &= \left(q\frac{\partial P}{\partial t} + P\frac{\partial q}{\partial t} - \frac{\partial TC}{\partial t}\right)dt + (MR - MC)dq \end{aligned} \tag{42}$$

This indicates that whatever the optimal rate of change of profits might be, it will not be an output level at which marginal revenue equals marginal cost. At this level of output, the contribution from the second term in the total differential would be zero, but since dq will normally be positive, marginal revenue should exceed marginal cost to provide some contribution from change in output to the change in profit.

This result doubtless appears paradoxical to economists who have been raised on the belief that equating marginal cost to marginal revenue maximises profit, but it is a straightforward result when one realises that the Marshallian condition is one for static profit maximisation, when "time is held constant" and only variations in quantity are considered. In that case, given the simple monotonic functions that are assumed for price and cost, profit has a single maximum, and locating the zero of its differential identifies this maximum. However, when one considers time as well, this local maximum with respect to quantity does not identify the path of change over time that maximises the rate of growth of profit.

Assuming that the dynamic goal involves maximising the rate of growth of profit over time, the dynamic objective of the firm can be said to be to maximise the differential of the log of profit:

```
In[9]:=  D@Log@Profit@q@tD, tDD, tD

Out[9]=  IP@q@tD, tD q¢@tD - TC^{H0,1L}@q@tD, tD +
           q@tD IP^{H0,1L}@q@tD, tD + q¢@tD P^{H1,0L}@q@tD, tDM - q¢@tD TC^{H1,0L}@q@tD, tDM •
         HP@q@tD, tD q@tD - TC@q@tD, tDL
```

(43)

Making the heroic assumption that this function has a single maximum, it occurs when the differential of this differential with respect to time equals zero:

```
In[10]:=  D@D@Log@Profit@q@tD, tDD, tD, tD Š 0

Out[10]=  - IP@q@tD, tD q¢@tD - TC^{H0,1L}@q@tD, tD + q@tD IP^{H0,1L}@q@tD, tD + q¢@tD P^{H1,0L}@q@tD, tDM -
              q¢@tD TC^{H1,0L}@q@tD, tDM^2 • HP@q@tD, tD q@tD - TC@q@tD, tDL^2 +
                        1
           ─────────────────────────────────
           P@q@tD, tD q@tD - TC@q@tD, tD
            IP@q@tD, tD q¢¢@tD - TC^{H0,2L}@q@tD, tD + 2 q¢@tD IP^{H0,1L}@q@tD, tD + q¢@tD P^{H1,0L}@q@tD, tDM -
              q¢¢@tD TC^{H1,0L}@q@tD, tD - q¢@tD TC^{H1,1L}@q@tD, tD +
              q@tD IP^{H0,2L}@q@tD, tD + q¢¢@tD P^{H1,0L}@q@tD, tD + q¢@tD P^{H1,1L}@q@tD, tD +
                 q¢@tD IP^{H1,1L}@q@tD, tD + q¢@tD P^{H2,0L}@q@tD, tDMM -
              q¢@tD ITC^{H1,1L}@q@tD, tD + q¢@tD TC^{H2,0L}@q@tD, tDMM == 0
```

(44)

It should be evident from the complexity of this expression, and the presence of many terms in it which cannot be known (such as the rate of change of quantity over time), that the values of marginal cost and marginal revenue are of no assistance to a firm in helping it to work out how to maximise the rate of growth of its profits over time.

3 Conclusion

There are three simple, intuitive explanations of why the conventional theory of the firm is erroneous.

Firstly, the conventional theory starts with independent market supply and demand functions. The demand function therefore has a marginal revenue function whose existence is completely independent of the conditions of supply: the marginal revenue curve exists regardless of how many firms there are in the industry, and its position cannot be altered by altering the number of producers. The intersection of this marginal revenue function with the marginal cost function—however derived—therefore shows the point at which aggregate profit in that industry is maximised. In arguing that a perfectly competitive industry produced past this point, the conventional theory in effect presumed a relationship between the conditions of supply and the conditions of demand, so that increasing the number of suppliers somehow made the demand curve and its derivative identical. Clearly this is impossible.

Secondly, the model of perfect competition argues that an industry with many independent profit maximising firms will produce past the point at which aggregate industry profit is maximised, and thus individually and collectively incur a loss on the output produced beyond the point at which marginal cost equals marginal revenue. In effect, the accepted theory argues that a single producing agent in this industry—a monopoly—will produce the profit maximising amount, but that a large number of agents will produce more than this.

But why? Why should a group of non-interacting rational agents collectively decide on a production level which results in less profit than a single agent? If it is rational for a single profit maximising agent to choose a particular output level, then given the same conditions of production, a group of rational profit maximisers should reach the same decision. Production of a quantity at which price equals marginal cost can only occur with irrational, non-profit-maximising behaviour. This fallacy has survived because of the erroneous belief amongst economists that the demand curve facing the individual competitive firm is horizontal.

Thirdly, the theory of production has continued to rest upon static foundations long after the reasons Marshall gave for doing this have ceased to apply.[1] Techniques for the analysis of dynamic processes are now much more advanced than they were in Marshall's day, and yet the analysis of the firm has not fully exploited these.

Clearly our paper calls into question Marshall's entire edifice of supply and demand analysis, and much else besides.

Since price cannot equal marginal cost unless firms deliberately produce substantially more than the profit maximising level, the well-known result that a supply curve cannot be derived for a monopoly generalises to any industry structure. It can no longer be argued that price is set by the intersection of supply and demand, but at best that (in non-interacting models of firm behaviour) output is determined by the intersection of marginal revenue and marginal cost, and the market price is set by the demand curve at this quantity.

The welfare comparisons between different industry structures (at least those of monopoly and perfect competition) are invalid. Diminishing marginal productivity makes definitive comparisons between different industry structures impossible, since cost functions must differ.

[1] "...dynamics includes statics... But the statical solution… is simpler...; it may afford useful preparation and training for the more difficult dynamical solution; and it may be the first step towards a provisional and partial solution in problems so complex that a complete dynamical solution is beyond our attainment." (Marshall, 1907 cited in Groenewegen 1996: 432)

Our paper also has implications for general equilibrium and other models which rely upon the assumption that price equals marginal cost, since this condition is clearly incompatible with profit maximising behaviour. Any model that makes this assumption must now justify it on the basis of interactions between agents which results in them producing more than the profit maximising level.

The conventional theory of the firm is therefore, to borrow a phrase from Kirman, an "empty citadel". What then should replace it?

Marshall once famously remarked that an economist should "(1) Use mathematics as a shorthand language, rather than as an engine of inquiry. (2) Keep to them till you have done. (3) Translate into English. (4) Then illustrate by examples that are important to real life. (5) Burn the mathematics. (6) If you can't succeed in 4, burn 3. This last I did often" (Marshall, cited in Pigou 1925: 427). Few economists have followed this advice, and in fact the profession has a history of ignoring empirical research—something akin to Marshall's step 4—which contradicts Marshall's model of the firm. We surmise that this failure to appreciate empirical data that did not conform to the accepted model was in part due to the belief that the mathematics behind the theory of the firm was incontrovertible. We hope we have demonstrated that this is not so. In this situation, the only sensible approach is to develop a theory of the firm which conforms to the substantial but neglected literature on the pricing and output behaviour of actual corporations.

Numerous researchers (Eiteman 1947 et seq., Haines 1948, Means 1972, Blinder et al. 1998—see Lee 1998 and Downward & Lee 2001 for surveys) have shown that the majority of actual firms have enormous fixed costs and constant or falling marginal costs, and determine their prices by a markup on variable costs. The size of the markup in turn is determined partly by the degree of competition (so that there is still some sense in which a more competitive industry is preferable to a less competitive one), but the size of the markup is also strongly motivated by the need to cover their fixed costs at a levels of output well within the current production capacity of the firm. Price is set by the firm prior to the market, and the firm attempts to sell as much of its output as it can at this price. Firms produce competing but heterogeneous products, and the main form of competition between firms is by product differentiation (by both marketing and R&D) rather than by price.

Means coined the term "the administered price thesis" for this perspective on the behaviour of the firm (Means 1972). Since that model accurately describes how actual firms behave, and since that behaviour cannot be reduced to profit maximisation via the equating of marginal cost and marginal revenue, economists should abandon Marshall and develop a microeconomics that is consonant with this empirical reality.

References

Aslambeigui, N, & Naples, M.I., (1997). "Scissors or horizon: Neoclassical debates about returns to scale, costs, and long run supply, 1926-1942", *Southern Economic Journal*, 64: 517-530.

Aumann, R.J., (1964). "Markets with a continuum of traders", *Econometrica* 32: 39-50.

Aumann, R.J., (1966). "Existence of competitive equilibria in markets with a continuum of traders", *Econometrica* 34: 1-17.

Bishop, R.L., (1948). "Cost discontinuities, declining costs and marginal analysis", *American Economic Review* 38: 607-617.

Blinder, A.S., Canetti, E., Lebow, D., & Rudd, J., (1998). *Asking About Prices: a New Approach to Understanding Price Stickiness*, Russell Sage Foundation, New York.

Clapham, J.H., (1922a). "Of empty economic boxes", *Economic Journal*, 32: 303-314.

Clapham, J.H., (1922b). "The economic boxes - a rejoinder", *Economic Journal*, 32: 560-563.

Conkling, R.L., (1999). "Marginal cost pricing for utilities: a digest of the California experience", *Contemporary Economic Policy*, 17: 20-32.

Downward, P. & Lee, F., (2001). "Post Keynesian pricing theory reconfirmed? A critical review of Asking About Prices", *Journal of Post Keynesian Economics* 23: 465-483.

Eiteman, W.J., (1947). "Factors determining the location of the least cost point", *American Economic Review* 37: 910-918.

Eiteman, W.J., (1948). "The least cost point, capacity and marginal analysis: a rejoinder", *American Economic Review* 38: 899-904.

Eiteman, W.J. & Guthrie, G.E., (1952). "The shape of the average cost curve", *American Economic Review* 42: 832-838.

Friedman, M., (1953). "The methodology of positive economics", in *Essays in Positive Economics*, University of Chicago Press, Chicago.

Groenewegen, P., (1995). *A soaring eagle: Alfred Marshall 1842-1924*, Edward Elgar, Cheltenham.

Haines, W.W., (1948). "Capacity production and the least cost point", *American Economic Review* 38: 617-624.

Harrod, R.F., (1934). "Doctrines of Imperfect Competition", *The Quarterly Journal of Economics*, 48: 442-470

Kalecki, M., 1937, "The principle of increasing risk", *Economica*, 440-447.

Keen, S., (2001). *Debunking Economics*, Pluto Press & Zed Books, Sydney.

Langlois, C., (1989). "Markup pricing versus marginalism: a controversy revisited", *Journal of Post Keynesian Economics* 12: 127-151.

Lee, F., (1996). "Pricing and the business enterprise", in C.W. Whalen (ed.), *Political Economy for the 21st century*, ME Sharpe, Armonk.

Lee, F., (1998). *Post Keynesian Price Theory*, Cambridge University Press, Cambridge.

Marshall, A., (1907 [1961]). *Principles of economics*, Macmillan, London.

Marshall, A. "Correspondence with Professor A. L. Bowley", in Pigou, A.C., (ed.), (1925). *Memorials of Alfred Marshall*, Macmillan Publishing Company, London.

Means, G.C, (1972). "The administered-price thesis reconsidered", *American Economic Review* 62: 292-306.

Pigou, A.C., (1922). "Empty economic boxes - a reply", *Economic Journal*, 36: 459-465.

Pigou, A.C., (1927). "The law of diminishing and increasing cost", *Economic Journal*, 41: 188-197

Pigou, A.C., (1928). "An analysis of supply", *Economic Journal*, 42: 238-257.

Robertson D.H., (1924). "Those empty boxes", *Economic Journal*, 34: 16-31.

Robertson D.H., (1930). "The trees of the forest", *Economic Journal*, 44: 80-89.

Robbins, L. (1928). "The representative firm", *Economic Journal*, 42: 387-404.

Sato, K., (1979). “A note on capital and output aggregation in a general equilibrium model of production”, *Econometrica* 47: 1559-1568.

Shepherd, W.G., (1984). “'Contestability' vs. competition”, *American Economic Review*, 74: 572-587.

Shove, G.F., (1930). “The representative firm and increasing returns”, *Economic Journal*, 44: 93-116.

Sraffa, P. (1926). “The Law of returns under competitive conditions”, *Economic Journal*, 40: 538-550.

Sraffa, P. (1930). “The trees of the forest—a criticism”, *Economic Journal*, 44: 89-92.

Stigler, G.J. (1957). “Perfect competition, historically considered”, *Journal of Political Economy*, 65: 1-17.