

Why economics textbooks must stop teaching the standard theory of the firm

Steve Keen, University of Western Sydney¹; s.keen@uws.edu.au

Abstract: The accepted theory of the firm abounds with fallacies, starting with one that has been known to be false since 1957—the horizontal demand curve. When these fallacies are corrected, nothing of substance remains. Price equals marginal cost is not a profit-maximizing equilibrium, competition does not lead to price equaling marginal cost, in general monopoly can be expected to generate a higher level of consumer welfare than competitive firms, and standard cost-curve aggregation of monopoly and perfect competition is only valid in highly restrictive circumstances. The accepted theory is internally incoherent and should no longer be taught to students of economics. We need a new microeconomics of the firm that should be based on empirical reality, rather than on superficially appealing but flawed *a priori* concepts. Substantial empirical research since 1920 has contradicted the 19th century *a priori* notion of firms producing homogeneous output under conditions of diminishing marginal productivity: real firms produce heterogeneous products under conditions of constant or falling marginal cost and set price well above marginal and average cost. This is the world we should model, and teach to our students.

Keywords: Microeconomics, competition, monopoly, perfect competition, oligopoly, profit maximization, Industrial Organization, Economics education

JEL Classifications: A20, C71, D20, D21, D41, D42, D43, D46, D50, D60, L11, L13, L21

Why economics textbooks must stop teaching the standard theory of the firm

When I wrote *Debunking Economics* (Keen 2000), I was aware that its most controversial part would be Chapter 4, “Size Does Matter”—and not just because of its title. The other chapters explained established critiques of conventional economic theory to a non-mathematical audience; Chapter 4 outlined a completely new and hitherto unpublished critique of one of the two keystones of introductory instruction in economics, the theory of the firm.

The reaction to this chapter has been as hostile as I expected. The consensus amongst economists has been that whatever the merits of the rest of the book, in Chapter 4 I simply didn’t know what I was talking about.

In one sense, that was true: I had only realized the basis to the critique while drafting the chapter, and I didn’t have time to fully explore its implications prior to publication.

I’ve since had the time, and the implications for the teaching of economics are profound—so profound that the best option is to abandon the theory of the firm altogether, and start teaching an empirically based alternative until an adequate new theory is produced. Before I explain why, I’ll repeat the key aspects of the standard canon.

1 The conventional canon

All undergraduate microeconomics textbooks teach students two strong welfare conclusions about competition:

- Perfectly competitive industries produce an aggregate quantity at which the market price in equilibrium equals the marginal cost of production, whereas a monopoly produces a lower quantity at which marginal cost equals marginal revenue. The level of output is lower and the price higher under monopoly than under perfect competition;
- Perfect competition results in the maximization of consumer surplus, whereas monopoly results in a transfer of some consumer surplus to the producer as well as a deadweight welfare loss.

In addition, some textbooks provide a reassuring link between the chimera of perfect competition and the real world, that profit-maximizing behavior converges to the perfectly competitive ideal as the number of firms in the industry rises. Using a formula first developed by Stigler (Stigler 1957), it is alleged that price rapidly converges to the perfectly competitive ideal of marginal cost as the number of firms in an industry rise. For example, with 50 firms and a market elasticity of demand of -2, market price will be within one per cent of the competitive ideal.

Figure 1 (taken from Mankiw 2004 Chapter 15) puts the first two standard arguments. According to Mankiw (and all other textbooks), the reason for this difference in behavior is not malice aforethought by the monopoly, but simply a difference in the nature of the demand curve perceived by it and competitive firms. The monopolist faces the entire industry demand curve, which is downward sloping ($\frac{dP}{dQ} < 0$), while each competitive firm is a “price-taker” who cannot influence the market price, so that the demand curve the perfectly competitive firm perceives is horizontal ($\frac{dP}{dq} = 0$) at the market price.

As a result, marginal revenue for the monopoly is less than the market price, while marginal revenue for a competitive firm equals the market price. For the monopolist, the mathematics is as in equation (1); since the monopolist maximizes profit by producing the quantity at which marginal cost equals marginal revenue, the monopoly price will exceed marginal cost:

$$MR_M = \frac{d}{dQ}(P \cdot Q) = P \cdot \frac{dQ}{dQ} + Q \cdot \frac{dP}{dQ} = P + Q \cdot \frac{dP}{dQ} < P \quad (1)$$

The mathematics for the competitive firms, on the other hand, is as in equation (2)

$$MR_{PC} = \frac{d}{dq}(P \cdot q) = P \cdot \frac{dq}{dq} + q \cdot \frac{dP}{dq} = P + q \cdot \frac{dP}{dq} = P + q \cdot 0 = P \quad (2)$$

Thus while competitive firms follow precisely the same profit maximizing guideline of equating marginal revenue and marginal cost, the proposition that the firm’s marginal revenue equals the market price means that each firm produces where marginal cost equals the market price. Therefore the rising portion of the marginal cost curve of the firm becomes its supply curve, and the sum of all firms’ supply curves equals the supply curve for the industry. On the other hand, it is not possible to

derive a “supply curve” for a monopoly, since there is a different marginal revenue curve for every demand curve, and the monopolist produces well above its marginal cost curve.

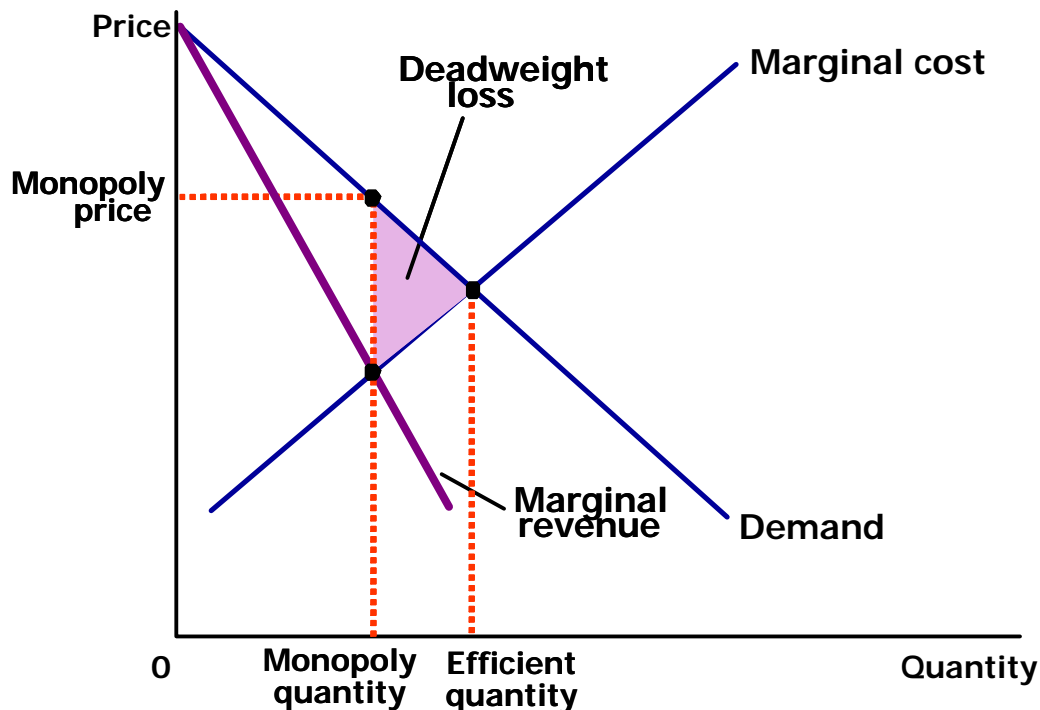


Figure 1: Standard welfare comparison of monopoly and perfect competition

At the aggregate market level, the intersection of the competitive industry supply curve with the market demand curve determines the equilibrium price. The supply curve represents the marginal cost of production of a commodity, while the demand curve represents the marginal benefit of its consumption. Where the two marginals are equal, the gap between total benefits and total costs is greatest. With the area above the price and below the demand curve representing consumer welfare, and the area below price and above the supply curve representing producer welfare, overall social welfare is maximized by perfect competition. On the other hand, with a monopoly supplier, there is a transfer of surplus from consumers to the producer, and a deadweight loss of welfare due to the monopoly.

Almost every proposition in this conventional textbook argument is false.

2 “Horizontal” demand curves

One of these—that the demand curve facing the individual competitive firm is horizontal—has been known to be false since 1957, when Stigler published the following simple piece of calculus in “Perfect competition, historically considered”:

$$\frac{dP}{dq_i} = \frac{dP}{dQ} \frac{dQ}{dq_i} = \frac{dP}{dQ} \quad (3)$$

Elaborating on Stigler’s spartan use of the Chain Rule, the slope of the demand curve facing the individual firm equals the slope of the market demand curve, multiplied by how much market output changes given a change in output by a single firm. Since we’re dealing with competitive, non-colluding firms, a change in output by one firm doesn’t elicit any *instantaneous* reaction by the others. Therefore the quantity “how much market output changes given a change in output by a single firm” is one. As a result, the slope of the individual firm’s demand curve is exactly the same as the slope of the market demand curve.

We can make the mathematics more explicit using summation notation. Assuming n identical firms, total output Q is the sum of the outputs of the n firms each producing q_i units. Therefore:

$$\begin{aligned} \frac{dP}{dq_i} &= \frac{dP}{dQ} \cdot \frac{dQ}{dq_i} \\ &= \frac{dP}{dQ} \cdot \frac{d}{dq_i} \sum_{j=1}^n q_j \\ &= \frac{dP}{dQ} \cdot \frac{d}{dq_i} (q_1 + q_2 + \dots + q_i + \dots + q_n) \\ &= \frac{dP}{dQ} \cdot \left(\frac{d}{dq_i} q_1 + \frac{d}{dq_i} q_2 + \dots + \frac{d}{dq_i} q_i + \dots + \frac{d}{dq_i} q_n \right) \\ &= \frac{dP}{dQ} \cdot (0 + 0 + \dots + 1 + \dots + 0) \\ &= \frac{dP}{dQ} \end{aligned} \quad (4)$$

The graphical intuition is shown in Figure 1, which shows a market demand curve for an industry with a large number of firms. Consider the slope of the demand curve between Q_1 to Q_2 consisting of a change of ΔQ in output and ΔP in price. If a single firm changes its output by δq , then market price will change by δP .² The slope of any tiny line segment $\frac{\delta P}{\delta q}$ is equivalent to the slope of the overall section $\frac{\Delta P}{\Delta Q}$. Thus the individual competitive firm’s demand curve has exactly the same negative slope as the market demand curve.

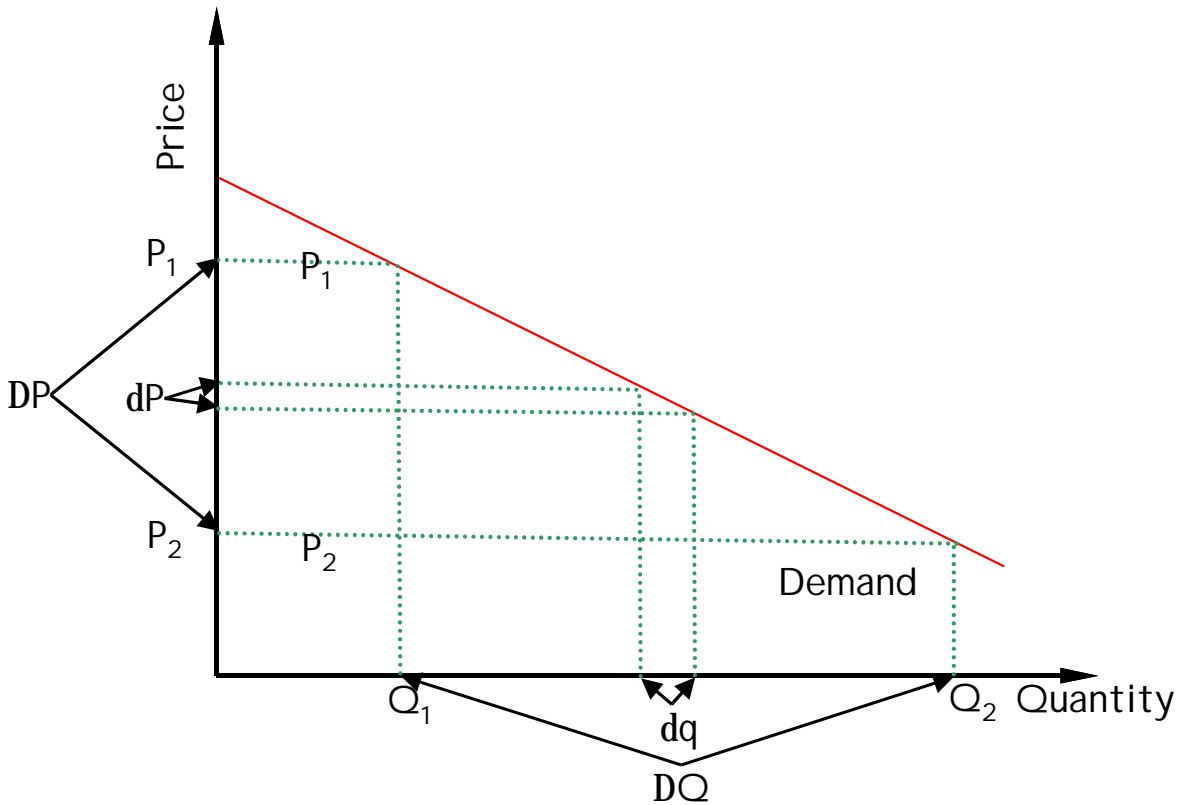


Figure 2: Slope of firm’s demand curve is identical to market demand curve

This argument can’t be avoided by an appeal to the proposition that competitive firms are “price takers”—so that marginal revenue equals price for competitive firms *by assumption*—because this simply introduces another logical contradiction into the theory, which can be illustrated using the concept of “conjectural variation”.³ Since competitive firms are independent, the amount that the i^{th} firm expects the rest of the industry Q_R to vary its output in response to a change in its output q_i is zero:

$$\frac{d}{dq_i} Q_R = 0 \tag{5}$$

The assumption that marginal revenue equals price for the i^{th} firm means that $\frac{d}{dq_i}(P \cdot q_i) = P$.

Now introduce $\frac{d}{dq_i} Q_R$ and $\frac{dP}{dq_i} = \frac{dP}{dQ} \frac{dQ}{dq_i}$ into this expression and expand:

$$\begin{aligned}
 \frac{d}{dq_i}(P \cdot q_i) &= P \cdot \frac{d}{dq_i}q_i + q_i \cdot \frac{d}{dq_i}P \\
 &= P + q_i \cdot \left(\frac{d}{dQ}P \cdot \frac{d}{dq_i}Q \right) \\
 &= P + q_i \cdot \left(\frac{d}{dQ}P \cdot \frac{d}{dq_i}(q_i + Q_R) \right) \\
 &= P + q_i \cdot \left(\frac{d}{dQ}P \cdot \left(\frac{d}{dq_i}q_i + \frac{d}{dq_i}Q_R \right) \right) \\
 &= P + q_i \cdot \left(\frac{d}{dQ}P \cdot \left(1 + \frac{d}{dq_i}Q_R \right) \right)
 \end{aligned} \tag{6}$$

Returning to our assumption that $\frac{d}{dq_i}(P \cdot q_i) = P$, the only way that this is possible is if $\frac{d}{dq_i}Q_R = -1$, but this contradicts (5): the concept of firm independence. This is “proof by contradiction” that the slope of the demand curve for the i^{th} firm must equal the slope of the market demand curve.

If the fact that the demand curve for the individual firm can't be horizontal has been in the literature for almost 50 years, why do textbook writers keep ignoring it? Partly because some comfort themselves with the observation that the elasticity of demand for a competitive firm $\left(e_i = \frac{P}{q_i} \cdot \frac{dq_i}{dP} \right)$ is so much larger than the elasticity of demand for the market as a whole $\left(E = \frac{P}{Q} \cdot \frac{dQ}{dP} \right)$. Sure, but this is a red herring. It has nothing to do with the relative slopes of the demand curves (which are identical), but is simply an artefact of the ratio between total industry output (Q) and the output of a single firm (q): $e_i/E = Q/q_i$. The elasticity of demand is also irrelevant to the calculation of marginal revenue: $\frac{d}{dq_i}P$ is the argument there, not $\frac{P}{q_i} \cdot \frac{dq_i}{dP}$ or any variant thereof.

Textbook writers may possibly ignore this problem because they know that, as well as pointing out a dilemma, Stigler also provided an alleged solution: that though marginal revenue exceeds price for all industry structures, it tends towards price as the number of firms in an industry rises. Assuming n identical firms each producing q units, Stigler introduced n and Q into the expression for the i^{th} firm's marginal revenue:

$$\begin{aligned}
 \frac{d}{dq_i}(P \cdot q_i) &= P + q \frac{dP}{dQ} \\
 &= P + \frac{Q}{n} \frac{P}{P} \frac{dP}{dQ} \\
 &= P + \frac{P}{n \cdot E}
 \end{aligned}
 \tag{7}$$

Stigler surmised that “this last term goes to zero as the number of sellers increases indefinitely” (Stigler 1957: 8), so that marginal revenue for the i^{th} firm converges to market price as the number of firms rises (this is true, but, as I’ll explain later, irrelevant). Perhaps knowledge of this expression comforts textbook writers that they can teach students a false proposition, because it doesn’t really matter: a more complex argument reaches the same conclusion anyway, and students can learn this approach later when they’re more familiar with economic reasoning.

Unfortunately, this salve to the conscience is also false, since—using Stigler’s Relation that $\frac{dP}{dq_i} = \frac{dP}{dQ}$ —it is easy to show that equating marginal cost and marginal revenue does not maximize profits.

3 Profit maximization formulae

“Maximize profits by equating marginal cost and marginal revenue” is one of the two key mantras of an undergraduate education in economics (the other being “maximize utility by equating relative prices and relative marginal utilities”). But what marginal revenue are we talking about: a change in revenue caused by the firm altering its own output level, or a change in revenue caused by what some or all of the other firms do? In a multi-firm industry, the i^{th} firm’s total revenue is a function not only of its own behavior, but also the behavior of all the other firms in the industry:

$$tr_i = tr_i \left(\sum_{j \neq i}^n q_j, q_i \right)
 \tag{8}$$

Once again defining Q_R as the output of the rest of the industry ($Q_R = \sum_{j \neq i}^n q_j$), a change in revenue for the i^{th} firm is properly defined as:

$$dtr_i(Q_R, q_i) = \left(\frac{\partial}{\partial Q_R} P(Q) \cdot q_i \right) dQ_R + \left(\frac{\partial}{\partial q_i} P(Q) \cdot q_i \right) dq_i \quad (9)$$

The accepted formula ignores the effect of the first term on the firm's profit (this isn't changes in the output of the rest of the industry *in direct response to a change in output by the i^{th} firm*, which is zero as discussed above $\left(\frac{d}{dq_i} Q_R = 0 \right)$, but simply changes that all other firms are making to output (dQ_R) as they independently search for a profit-maximizing level of output).

It is therefore obvious that the quantities produced will *not* be profit-maximizing if all firms apply the standard formula. However it is possible to work out a general profit-maximization formula for the single firm by first establishing the aggregate industry output level that would result if each firm in the industry did equate its marginal cost to its own-output marginal revenue.

In the following derivation I use Stigler's Identity $\frac{dP}{dq_i} = \frac{dP}{dQ}$, and the simple rule for the aggregation of marginal costs (or rather the horizontal summation of the marginal cost curves of firms in an industry) that $mc_i(q_i) = MC(Q)$.⁴

If all firms in an industry equate their own-output marginal revenue to marginal cost, then the sum of all these zeros is also zero. Expanding using summation notation, we have:

$$\begin{aligned} \sum_{i=1}^n \left(\frac{d}{dq_i} (P(Q) \times q_i - TC_i(q_i)) \right) &= \sum_{i=1}^n \left(P(Q) + q_i \frac{d}{dq_i} P(Q) \right) - \sum_{i=1}^n \left(\frac{d}{dq_i} TC_i(q_i) \right) \\ &= nP(Q) + \sum_{i=1}^n \left(q_i \frac{d}{dQ} P(Q) \right) - \sum_{i=1}^n mc(q) \\ &= nP(Q) + \frac{d}{dQ} P(Q) \sum_{i=1}^n q_i - \sum_{i=1}^n MC(Q) \\ &= nP(Q) + Q \frac{d}{dQ} P(Q) - n \cdot MC(Q) \\ &= (n-1)P(Q) + \left(P(Q) + Q \frac{d}{dQ} P \right) - n \cdot MC(Q) \\ &= (n-1)P(Q) + MR(Q) - n \cdot MC(Q) \\ &= 0 \end{aligned} \quad (10)$$

Equation (10) can be rearranged to yield:

$$MR(Q) - MC(Q) = -(n-1)(P(Q) - MC(Q)) \quad (11)$$

Since $n-1$ exceeds 1 in all industry structures except monopoly, and price exceeds marginal cost,⁵ the RHS of (11) is negative. Thus industry marginal cost *exceeds* marginal revenue if each firm equates its own-output marginal revenue to marginal cost, so that part of industry output is produced at a loss. These losses at the aggregate model must be born by firms within the industry, so that firms that equate their own-output marginal revenue to marginal cost are producing part of their output at a loss.

Equation (11) can be used to derive the actual profit-maximizing quantity for the industry and the firm:

$$\sum_{i=1}^n \{mr_i(q_i) - mc(q) - \frac{n-1}{n} \cdot (P(Q) - MC)\} = MR(Q) - MC \quad (12)$$

Setting this to zero identifies both the industry-level output Q_K and firm level output q_k that maximize profits, and the individual profit-maximizing strategy:

$$mr_i(q_k) - mc(q_k) = \frac{n-1}{n} \cdot (P(Q_K) - MC(Q_K)) \quad (13)$$

This formula obviously corresponds to the accepted formula for a monopoly. However for a multi-firm industry, (13) indicates that firms maximize profits, not by equating their own-output marginal revenue and marginal cost, but by producing where their own-output marginal revenue *exceeds* their marginal cost.

So a principle that, if you are a microeconomics teacher, you have taught to thousands of students, is false: equating marginal cost and marginal revenue *does not* maximize profits. Instead, the true profit maximization rule is as shown in Figure 3: firms maximize profits by making the gap between own-output marginal revenue and marginal cost equal to $(n-1)/n$ times the gap between price and marginal cost. The intersection of two curves is not “where the action is” in this instance.

Once we plug in different industry structures into this true profit maximization formula, it turns out that there is no difference between competitive industries and a monopoly: if they have the same cost functions, then they will produce exactly the same amount and sell at the same price.

4 Perfect competition equals monopoly

Consider a linear demand curve $P(Q) = a - b \cdot Q$ and n firms facing the identical marginal cost function $mc(q) = c + d \cdot q$.⁶ Using Stigler's Relation that $\frac{dP}{dq_i} = \frac{dP}{dQ}$, marginal revenue for the i^{th} such firm is:

$$\begin{aligned} mr_i &= P + q_i \cdot \frac{dP}{dQ} \\ &= a - b \cdot Q - b \cdot q_i \end{aligned} \quad (14)$$

Feeding this into the accepted formula yields:

$$\begin{aligned} a - b \cdot n \cdot q - b \cdot q - (c + d \cdot q) &= 0 \\ q &= \frac{a - c}{(n + 1) \cdot b + d} \end{aligned} \quad (15)$$

Aggregate output is therefore a function of n : $Q = n \cdot \frac{a - c}{(n + 1) \cdot b + d}$. This converges to the perfect competition ideal $\frac{a - c}{b}$ as $n \rightarrow \infty$: $\lim_{n \rightarrow \infty} n \cdot \frac{a - c}{(n + 1) \cdot b + d} = \frac{a - c}{b}$.

Feeding Stigler's Relation into the correct profit maximizing formula (13) yields:

$$\begin{aligned} P - bq - (c + d \cdot q) &= \frac{n - 1}{n} (P - (c + d \cdot q)) \\ q &= \frac{a - c}{2 \cdot n \cdot b + d} \end{aligned} \quad (16)$$

Aggregate output is therefore:

$$Q = n \cdot \frac{a - c}{2 \cdot n \cdot b + d} \quad (17)$$

If $n = 1$, this coincides with the accepted formula's prediction for a monopoly. But as $n \rightarrow \infty$ the limiting output is precisely half that predicted by the conventional formula: $\lim_{n \rightarrow \infty} n \cdot \frac{a - c}{2 \cdot n \cdot b + d} = \frac{a - c}{2 \cdot b}$

It is also easily shown that the revised formula results in a much larger profit for each individual firm in the industry than the accepted "profit maximizing" formula. Profit is total revenue minus total cost, where total cost is the integral of marginal cost. Using k for fixed costs, profit at output q is:

$$\begin{aligned} \pi(q) &= P(Q) \cdot q - tc(q) \\ &= (a - b \cdot n \cdot q) \cdot q - \left(k + c \cdot q + \frac{1}{2} \cdot d \cdot q^2 \right) \end{aligned} \quad (18)$$

Feeding in the accepted formula's quantity q_c yields a profit level of:

$$\pi(q_c) = \left(\frac{1}{2} \frac{(a-c)^2 \cdot (2 \cdot b + d)}{((n+1) \cdot b + d)^2} \right) - k \quad (19)$$

The revised formula's quantity q_k yields a profit level of:

$$\pi(q_k) = \left(\frac{1}{2} \frac{(a-c)^2}{2 \cdot b \cdot n + d} \right) - k \quad (20)$$

The revised formula's profit level equals the accepted formula's for a monopoly where $n=1$ but exceeds it for $n>1$:

$$\pi(q_k) - \pi(q_c) = \frac{1}{2} \cdot b^2 \cdot \frac{(n-1)^2 \cdot (a-c)^2}{(2 \cdot b \cdot n + d) \cdot ((n+1) \cdot b + d)^2} \quad (21)$$

A numerical example indicates just how substantially the accepted formula's level of output *exceeds* the profit-maximizing level for an individual firm. With 100 firms and the parameters $a = 100$, $b = \frac{1}{1000000}$, $c = 20$, $d = \frac{1}{100000}$, the conventional formula results in an output level of 720 thousand units and a profit of \$3.1 million per firm. The revised formula results in an output level of 381 thousand units and a profit of \$15.2 million per firm. Figure 3 illustrates the difference in profits as a function of the number of firms.

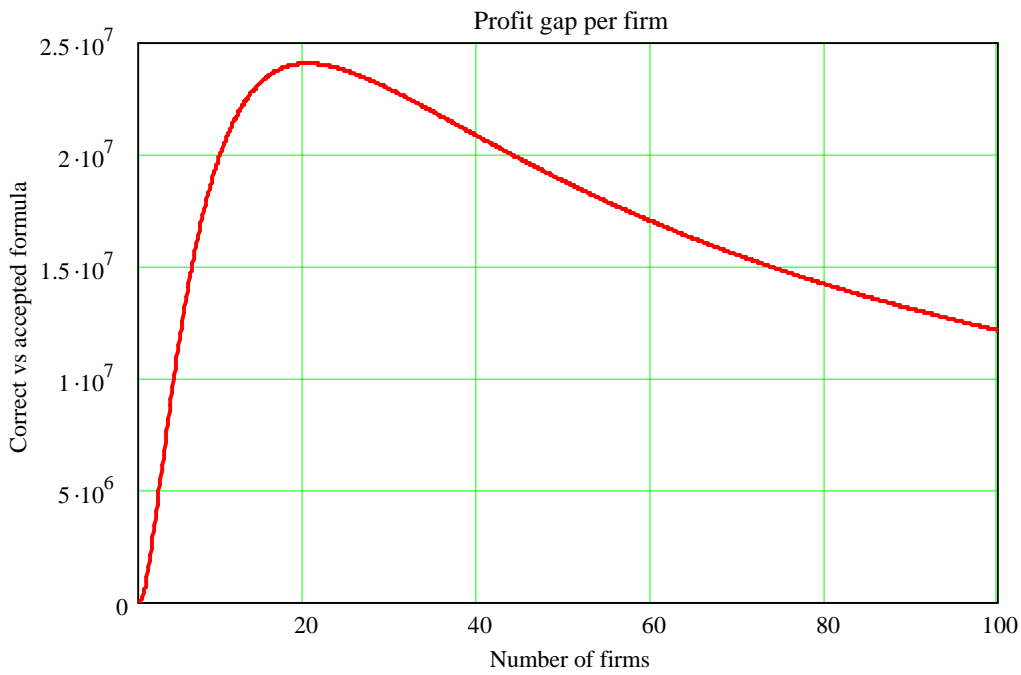
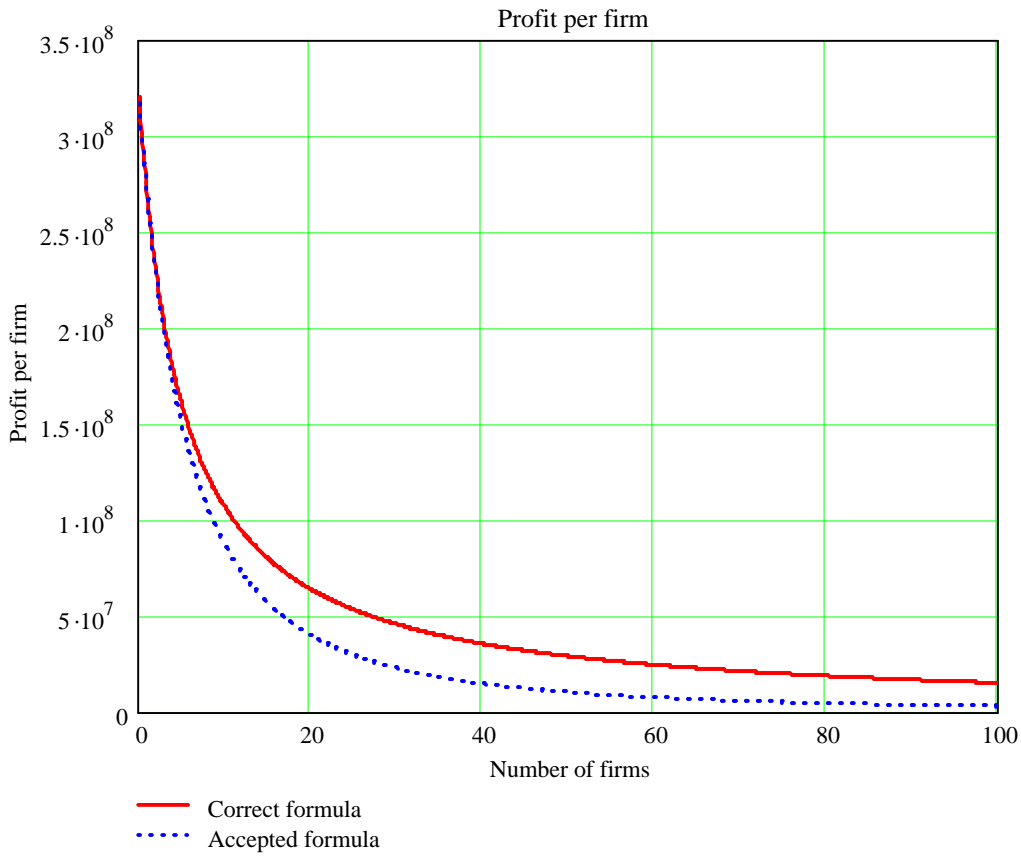


Figure 3: Profit gap between correct and accepted formula

Therefore profit-maximizing competitive firms will in the aggregate produce the same output as a monopoly, and sell it at the same price (given the same cost function, of which more later). Both

market structures have identical welfare implications, and the deadweight loss that has been previously attributed solely to monopoly behavior is in fact due to profit maximizing behavior.

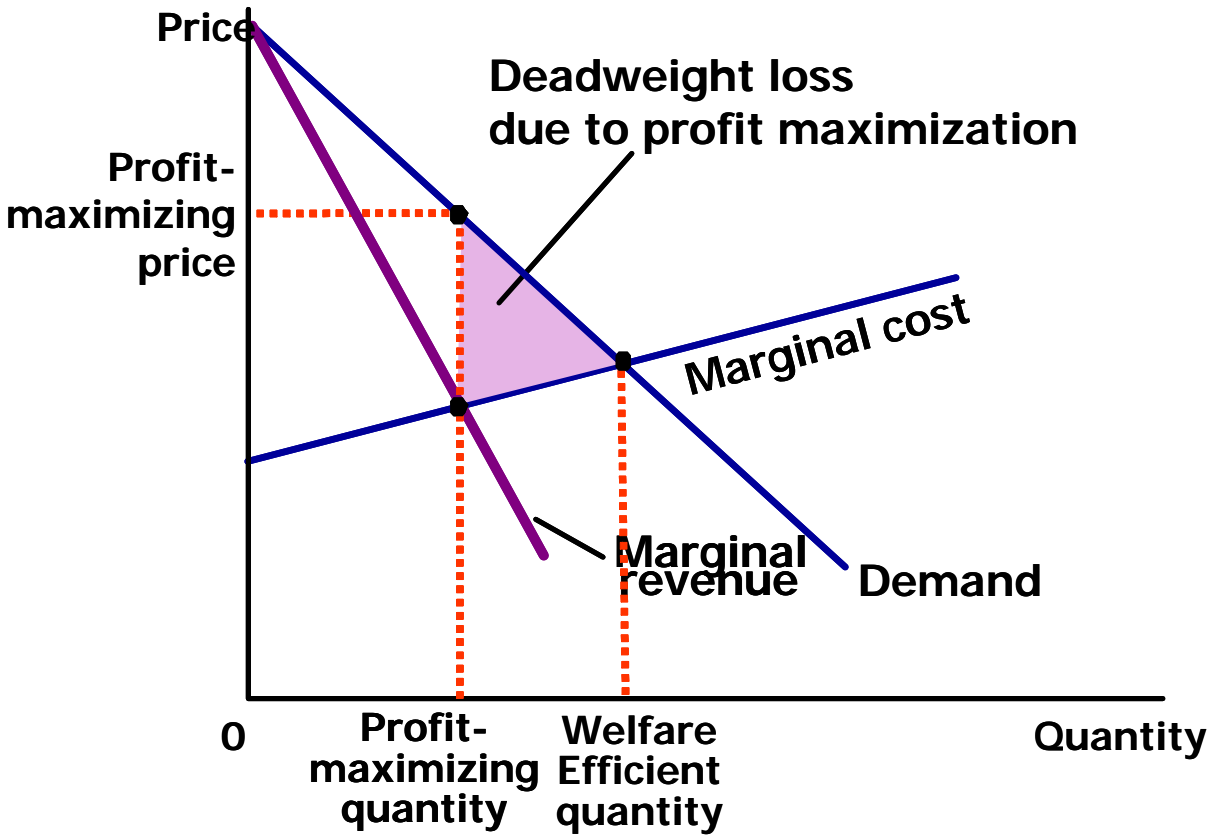


Figure 4: Output and welfare, exceptional welfare comparable case

This accurate profit-maximization rule results in the competitive firms producing at the same level as the monopoly, regardless of the number of firms in the industry.

5 Stigler's relation

This is why I described Stigler's relation $MR_i = P + \frac{P}{n \cdot E}$ as correct but irrelevant: while the i^{th} firm's own-output marginal revenue will converge to the market price as the number of firms in the industry rises, the market price to which convergence occurs is the "monopoly" price. If we feed Stigler's Relation into the true profit maximization formula and solve for market price, we get:

$$P + \frac{P}{n \cdot E} - MC = \frac{n-1}{n}(P - MC) \tag{22}$$

$$\left(1 + \frac{1}{E}\right) \cdot P = MC$$

As is well-known, the LHS of (22) is aggregate industry marginal revenue: $MR = \left(1 + \frac{1}{E}\right) \cdot P$. Price in a competitive industry with profit-maximizing firms therefore converges to the "monopoly"

price where aggregate marginal revenue equals aggregate marginal cost, regardless of the number of firms in the industry.

6 Monopoly is better...?

The above example assumes that the output levels of monopoly and competition can be compared.⁷

In fact, this comparison can only be made when the aggregate cost curves of the two industry structures are identical. Though economics textbooks draw this as the standard situation, in fact it can apply only in 3 restrictive cases: where the monopoly comes into being by taking over and operating all the firms of the competitive market; where the monopoly and the competitive firms operate under conditions of constant identical marginal cost; or where the monopoly and competitive firms have differing marginal costs that happen to result in equivalent aggregate marginal cost functions.

Consider an n -firm competitive industry and an m -plant monopoly. For the aggregate marginal cost curves to coincide, then the horizontal aggregation of the quantities produced at each level of marginal cost must sum to the same aggregate marginal cost function. This gives us an aggregation condition that that $Q=n \cdot q_c=m \cdot q_m$ where q_c is the output of a competitive firm and q_m is the output of a monopoly plant at the same marginal cost level.

In general we might draw a diagram like Figure 5:

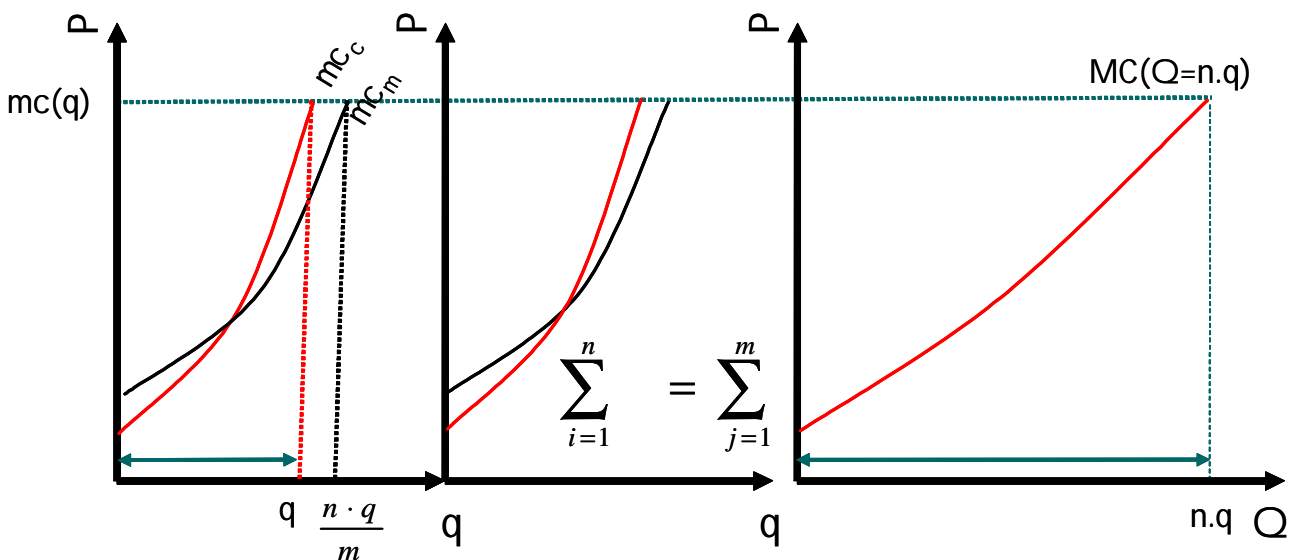


Figure 5: Horizontal aggregation of quantities from n firms or m plants

If $m=n$ —if the monopoly simply takes over all the competitive firms—then the two curves mc_c and mc_m coincide and any cost function will work. If however $m < n$ (the monopoly has less plants and operates on a larger scale than the competitive firms), then Figure 5 has to be amended in two ways: firstly, the curves can't intersect, since at that level $m \cdot q < n \cdot q$ and the aggregate curve couldn't be drawn; secondly and for the same reason, the y-intercept must be the same. This gives us Figure 6:

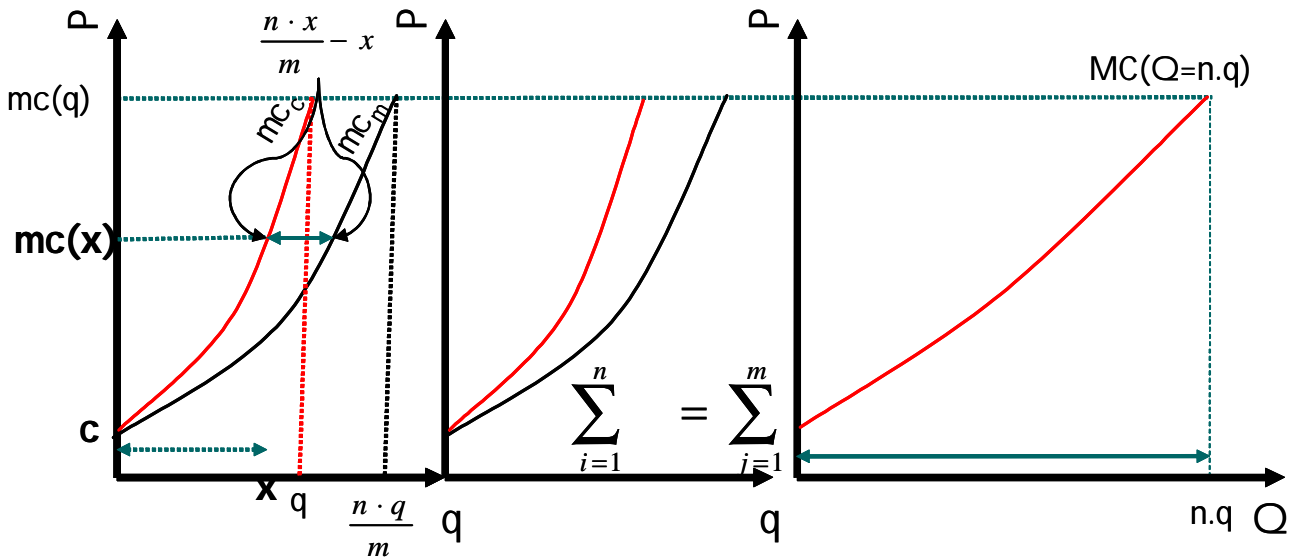


Figure 6: Amended horizontal aggregation of quantities from n firms or m plants

Again, the condition that $Q = n \cdot q_c = m \cdot q_m$ shows that this figure must be amended. The maximal marginal cost level shown of $mc(q)$ results in each competitive firm producing q units of output and each monopoly plant producing $\frac{n}{m} \cdot q$ units (so that the aggregate output in both cases is $Q = n \cdot q$). The same condition applies at any intermediate level of marginal cost $mc(x)$, as shown in Figure 6: the monopoly's plants must produce $\frac{n}{m} \cdot x$ units of output at a marginal cost level that results from x units of output from each competitive firm so that in the aggregate $Q(x) = n \cdot x = m \cdot \frac{n}{m} x$. This is only possible if mc_c and mc_m are straight-line functions, where the slope of mc_m is $\frac{m}{n}$ times the slope of mc_c . This is illustrated in Figure 7:

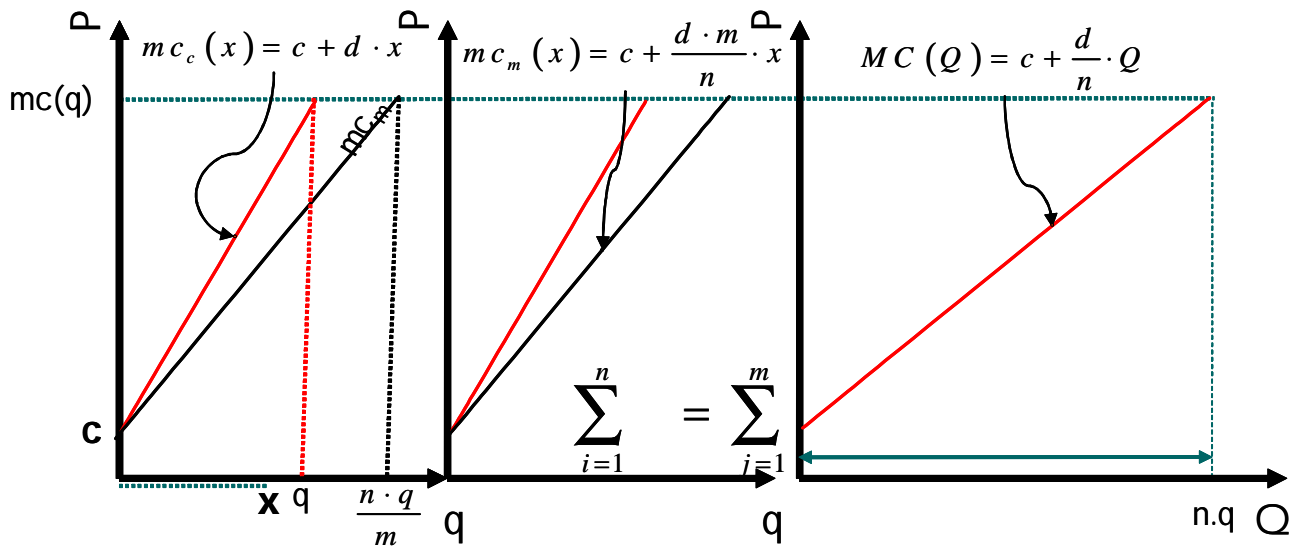


Figure 7: Horizontal aggregation condition for identical aggregate marginal cost

With any more general nonlinear marginal cost function—or even linear marginal cost functions where the slopes don't have this correspondence—the aggregate marginal cost curve for a competitive industry *must* differ from that for a monopoly. We know from the above analysis that both will set price where aggregate marginal revenue equals marginal cost. Therefore whichever market structure has the lower marginal costs will produce the greater output. So theory alone can't decide which is better—economists have to get their hands dirty and actually do real empirical work.

What is such work likely to find? In general, the odds are that monopolies (or the larger plants that tend to accompany more concentrated industry structures) will have *lower* marginal costs than competitive firms. Rosput (1993) gives an instructive illustration (in the case of gas delivery) of how greater economies of scale can result in lower marginal costs, even with the same technology:

“Simply stated, the necessary first investment in infrastructure is the construction of the pipeline itself. Thereafter, additional units of throughput can be economically added through the use of horsepower to compress the gas up to a certain point where the losses associated with the compression make the installation of additional pipe more economical than the use of additional horsepower of compression. The loss of energy is, of course, a function of, among other things, the

diameter of the pipe. Thus, at the outset, the selection of pipe diameter is a critical ingredient in determining the economics of future expansions of the installed pipe: *the larger the diameter, the more efficient are the future additions of capacity and hence the lower the marginal costs of future units of output.*” (Rosput 1993: 288; emphasis added)⁸

The general rule of Adam Smith’s pin factory also comes into play: the *specialization* that a larger scale of operation allows will lower marginal costs. In Ma & Pa Kettle’s corner shop, Ma and Pa have to accept deliveries, stock shelves, assist customers, make sales, do the accounts, order stock, etc. With Wal-Mart and its like, all these processes are specialized with both personnel and equipment, leading to higher output per worker and thus lower marginal costs.

With lower marginal costs, the monopoly will produce a greater quantity than the competitive industry and sell it at a lower price, resulting in a higher level of consumer surplus, as shown in Figure 8. Whereas this was a possibility under accepted theory, it is a *certainty* with any marginal cost difference in the monopoly’s favor when the theory is adjusted to eliminate aggregation errors.

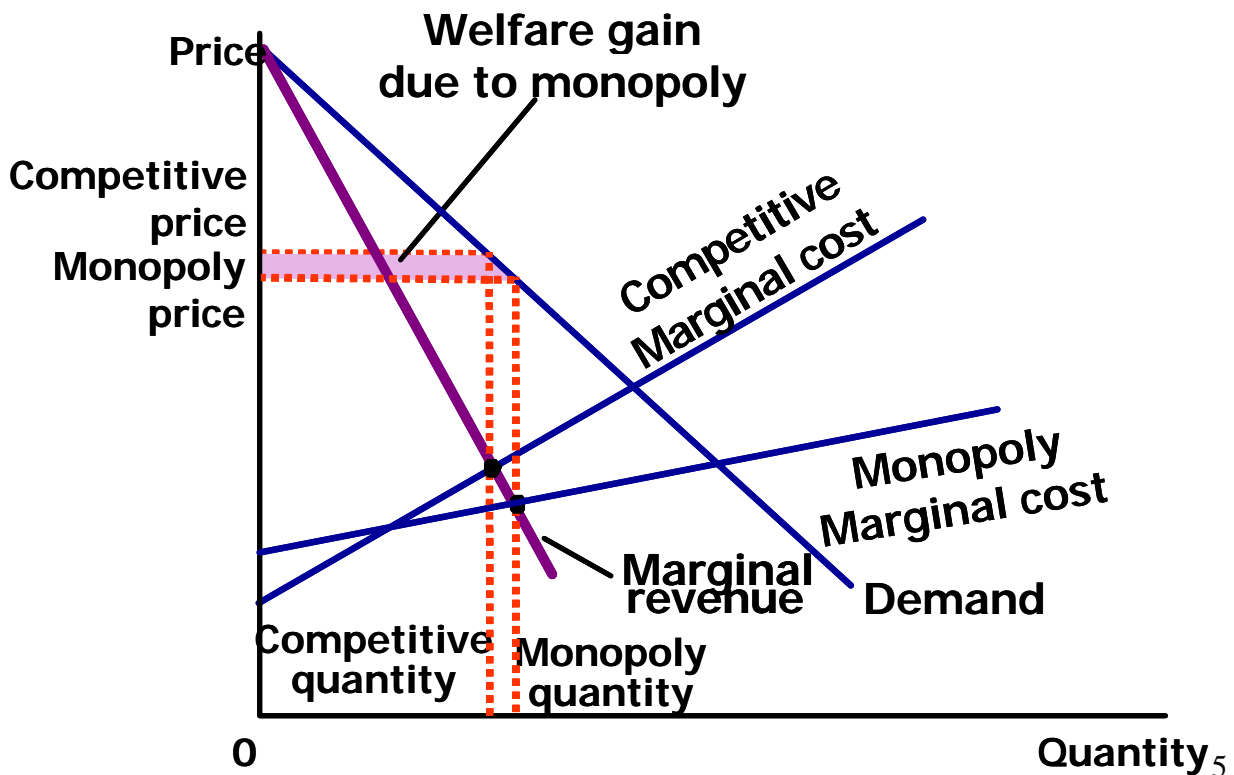


Figure 8: Welfare gain due to monopoly

7 Price equals marginal cost is not an equilibrium

The concept that everything happens in “equilibrium” is a peculiar obsession of economists that I derided in *Debunking Economics*: other sciences are quite content to model processes that normally occur out of equilibrium. So here’s a final conundrum for conventional theory: the “welfare ideal” of price equal to marginal cost is not an equilibrium, whereas the output level specified by the aggregation error amended formula is.

In general, aggregate profit is

$$\Pi(Q) = P(Q) \cdot Q - TC(Q) \tag{23}$$

Taking the amended formula first and using Q_K to signify the output level at which aggregate marginal cost equals aggregate marginal revenue, a small change in output δq results in an aggregate profit level of

$$\Pi(Q_K + \delta q) = P(Q_K + \delta q) \cdot (Q_K + \delta q) - TC(Q_K + \delta q) \tag{24}$$

Applying a Taylor’s series expansion to this, the new profit level is approximately

$$\begin{aligned}\Pi(Q_K + \delta q) &\approx \left(P(Q_K) + \delta q \cdot \frac{dP}{dQ} \right) \cdot (Q_K + \delta q) - \left(TC(Q_K) + \delta q \cdot \frac{dTC}{dQ} \right) \\ &= P(Q_K) \cdot Q_K + P(Q_K) \cdot \delta q + \delta q \cdot \frac{dP}{dQ} \cdot Q_K + \delta q \cdot \frac{dP}{dQ} \cdot \delta q - TC(Q_K) - \delta q \cdot \frac{dTC}{dQ}\end{aligned}\quad (25)$$

This expansion contains $P(Q_K) \cdot Q_K - TC(Q_K) = \Pi(Q_K)$ which we can now subtract from both sides to yield

$$\Pi(Q_K + \delta q) - \Pi(Q_K) \approx \left(P(Q_K) + \frac{dP}{dQ} \cdot Q_K \right) \cdot \delta q + \delta q^2 \cdot \frac{dP}{dQ} - \delta q \cdot \frac{dTC}{dQ}\quad (26)$$

Since the output level Q_K is that at which aggregate marginal revenue $P(Q_K) + \frac{dP}{dQ} \cdot Q_K$ equals aggregate marginal cost $\frac{dTC}{dQ}$, the first term on the RHS of (26) $\left(P(Q_K) + \frac{dP}{dQ} \cdot Q_K \right) \cdot \delta q$ cancels with the third $(\delta q \cdot \frac{dTC}{dQ})$ leaving

$$\Pi(Q_K + \delta q) - \Pi(Q_K) \approx \delta q^2 \cdot \frac{dP}{dQ}\quad (27)$$

Since $\frac{dP}{dQ} < 0$, this is necessarily negative: hence any change in output will reduce profit.⁹ The aggregate output level Q_K is thus a profit-maximizing equilibrium.

Using Q_{PC} for the convergence point of the standard formula, equation (29) becomes

$$\Pi(Q_{PC} + \delta q) - \Pi(Q_{PC}) \approx P(Q_{PC}) \cdot \delta q + \delta q \cdot \frac{dP}{dQ} \cdot Q_{PC} + \delta q \cdot \frac{dP}{dQ} \cdot \delta q - \delta q \cdot \frac{dTC}{dQ}\quad (28)$$

Since this is the output level at which price equals marginal cost, we have $P(Q_{PC}) = \frac{dTC}{dQ}$. We cancel the first and last terms $(P(Q_{PC}) \cdot \delta q$ and $\delta q \cdot \frac{dTC}{dQ})$ to give us the residual:

$$\Pi(Q_{PC} + \delta q) - \Pi(Q_{PC}) \approx \frac{dP}{dQ} \cdot (Q_{PC} + \delta q) \cdot \delta q\quad (29)$$

$\frac{dP}{dQ}$ is negative and $(Q_{PC} + \delta q)$ is positive, so the sign of $\Pi(Q_{PC} + \delta q) - \Pi(Q_{PC})$ is the inverse of the sign of δq . If $\delta q > 0$, profit will fall; if $\delta q < 0$, profit will rise. Thus any firm that decreases its output from the level at which its marginal cost equals marginal price will increase its profits, and the increase in profit will also to some extent be transmitted to all other firms via an increase in market price. The output level at which price equals marginal cost is therefore mathematically not a profit-maximizing equilibrium.

What if we assume “price-taking behavior” (so that $\frac{dP}{dq} = 0$ in equation 34), rather than using Stigler’s relation that $\frac{dP}{dq} = \frac{dP}{dQ}$? Then

$$\Pi(Q_{PC} + \delta q) - \Pi(Q_{PC}) \approx P(Q_{PC}) \cdot \delta q + 0 + 0 - \delta q \cdot \frac{dTC}{dQ} \quad (30)$$

Since this is the output level at which price equals marginal cost, the first and fourth terms cancel. Then we get that $\Pi(Q_{PC} + \delta q) - \Pi(Q_{PC}) \approx 0$; changing output from the “profit maximizing” level doesn’t change profits! This is a logical contradiction—thus the assumption that $\frac{dP}{dq} = 0$ must be false, as illustrated earlier using conjectural variation. This is another way of confirming that Stigler’s Relation $\frac{dP}{dq} = \frac{dP}{dQ}$ is right and the conventional textbook argument that $\frac{dP}{dq} = 0$ is wrong.

8 Conclusion

What appears to be a coherent theory of profit maximizing firms is in fact an incoherent shambles. A key assumption in the textbook argument is false, has been known to be false since 1957(!), and leads to logical contradictions if it is maintained. The hallmark of the theory, the conclusion that competition is welfare superior to monopoly, is invalid: given profit maximizing behavior, competitive industries are at best welfare identical to monopoly. Given the economies of scale, specialization and research and development advantages of large firms over small ones, in most cases monopoly (or other structures characterized by large firms with large market shares) will be superior to competition. And equating marginal revenue and marginal cost doesn’t maximize profits (except for a monopoly). The iconic graph of a firm maximizing profits by producing where the marginal revenue and marginal cost curves intersect has to be amended to show the firm maximizing profits at an output level that results in a *gap* between marginal revenue and marginal cost (see Figure 9).

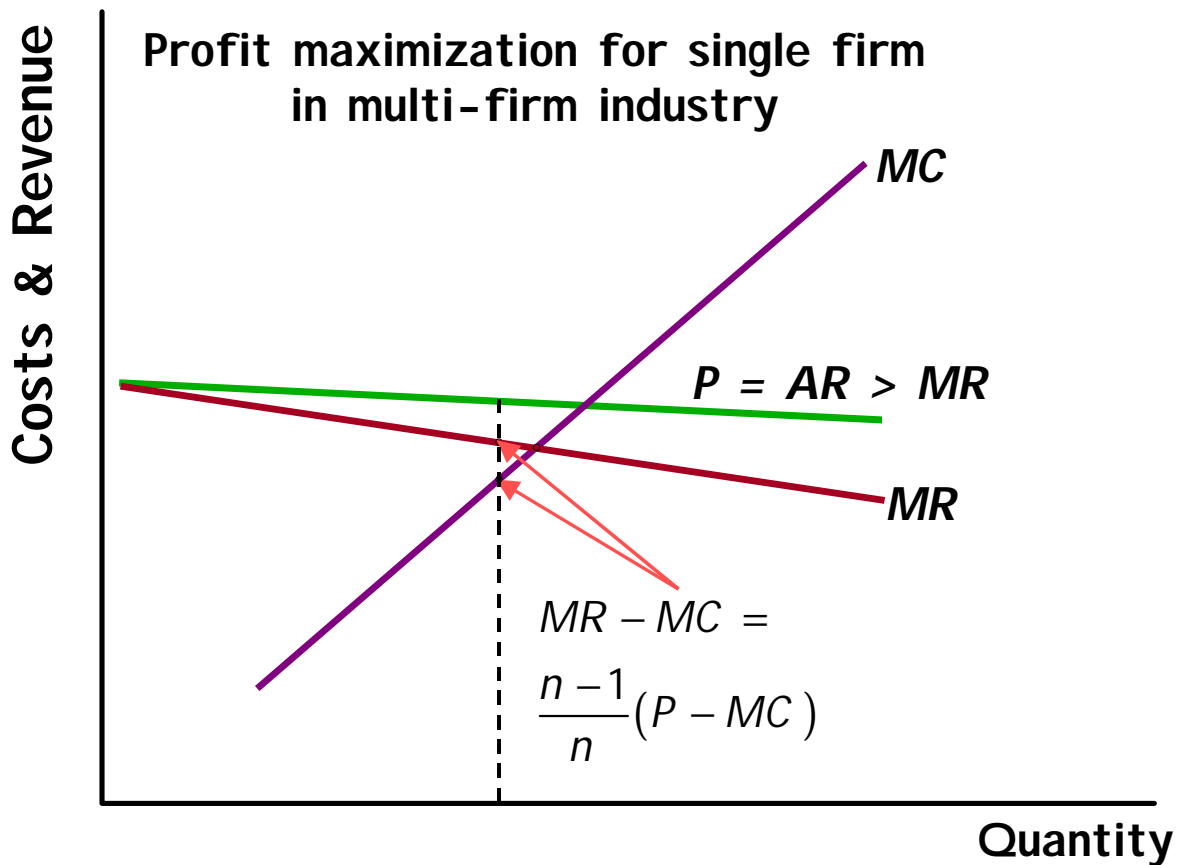


Figure 9: The general profit-maximizing rule in a multi-firm industry

So what should teachers of economics do? Continue teaching this “theory” because, to borrow from Maggie Thatcher, “There is NO alternative?” Stick with the theory but now argue that monopoly is good and perfect competition bad?

No! We should instead confront the reality that the *a priori* analysis that gave us this theory—starting from the concept of diminishing marginal productivity on one hand and market-clearing, competitive equilibrium prices on the other—has led us into a blind alley. So if *a priori* reasoning led us up the garden path, *let’s be radical and see what actually happens in the real world.*¹⁰

Fortunately, many researchers have already done this (about 120 studies have been published at last count), and the news is good and bad.

The good news is that the researchers all found the same result, so there is a clear and unambiguous reality to be modelled. The bad, from the perspective of conventional textbook

teaching, is that what they found seriously contradicts the *a priori* concepts of diminishing marginal productivity and market-clearing, competitive equilibrium prices.¹¹ These studies (Eiteman 1947 et seq., Haines 1948, Means 1972, Blinder et al. 1998—see Lee 1998 and Downward & Lee 2001 for surveys) reveal that the vast majority of actual firms have production functions characterized by large average fixed costs and *constant or falling* marginal costs. This cost structure, which is described as “natural monopoly” in economic literature and portrayed as an exception to the rule of rising marginal cost, actually appears to be the empirical reality for 95 per cent or more of real firms and products (Becker 2004 also points out that the “new economy” firms have near-zero constant marginal costs, but price substantially above this level). Once a breakeven sales level has been achieved, each new unit sold adds significantly to profit, and this continues out to the last unit sold—so that marginal revenue is always significantly above marginal cost (more so even than in the mathematical analysis above).

Firms compete not so much on price but on heterogeneity: Competitors’ products differ qualitatively, and the main form of competition between firms is not price but product differentiation (by both marketing and R&D).

The prices that rule in real markets also are not the market clearing prices of our defunct theory, but what Means termed “administered prices”, set largely by a markup on variable costs, with the size of the markup reflecting many factors (of which the degree of competition is one).¹²

So teachers of economics aren’t entirely in the dark if they abandon the shoddy horse of marginal analysis. We have substantial empirical data on how firms actually behave; we also have substantial empirical data on what this means at the market level, with numerous price and quantity series, and substantial research into marketing and management practices by our colleagues in neighbouring academic departments.

What we lack is a theory to meld this observed firm level behavior with the observed behavior of markets (though the much maligned Post Keynesian school of thought offers some useful guidance

here; see Lee 1998). But that is what economists are supposed to do: produce theories which explain and interpret the empirical economic reality.

One clue as to what such a theory might be comes simply from lifting the veil that static reasoning has pulled over our eyes for the last century. The concept of an equilibrium between supply and demand is an abstraction from the reality that most markets are growing most of the time. If we confront this reality, then a first approximation is that firms should aim for an equilibrium between the *rate of growth* of supply and the *rate of growth* of demand. Expressing this as a differential equation, we get the objective:

$$\frac{d}{dt}S(t) - \frac{d}{dt}D(t) = 0 \quad (31)$$

Integrating this gives us

$$S(t) - D(t) = K \quad (32)$$

A constant of integration drops out of the equation: firms should attempt to keep the *gap* between supply and demand constant (this “gap” in the real world is the stock of unsold goods that virtually all firms maintain). Goodbye market-clearing prices—and perhaps economists should develop a theory of pricing based on an adjustment mechanism for stocks: price rises if stocks are falling, and vice versa.

Of course, the real world mechanism has to be more complex: in a growing economy, maybe firms would aim to keep their stocks growing at the same (or a higher) rate than the economy itself (so perhaps K is a function of time rather than a constant).

This should be the stuff of economics: finding out what actually happens in the real economy and finding a way to model it, and extend our understanding of economic reality as a result. We have to encourage our students to behave this way, because it is they who will probably undertake the task. We may not be able to help them all that much in it, but we can certainly try not to obstruct them by filling their minds with erroneous and unworkable concepts.

References

- Arrow, K.J. 1959. Towards a theory of price adjustments". In Abramovitz, M. et al., (eds.), *The Allocation of Economic Resources*, Stanford: Stanford University Press.
- 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.
- Becker, W.E. 2004. Good-Bye Old, Hello New in Teaching Economics. *Australian Journal of Economic Education* (forthcoming).
- 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*, New York: Russell Sage Foundation.
- Boland, L. A. 1981. On the Futility of Criticizing the Neoclassical Maximization Hypothesis. *American Economic Review* 71: 1031-1036.
- Conkling, R.L. 1999. Marginal cost pricing for utilities: a digest of the California experience. *Contemporary Economic Policy* 17: 20-32.
- Cournot, A. 1838[1988]. Of the competition of producers. In *Researches into the Mathematical Principles of the Theory of Wealth*, L. Hachette, Paris, reprinted in Daughety, A.F., (ed.), 1988. *Cournot Oligopoly*, Cambridge University Press, Cambridge.
- Downward, P. & Lee, F. 2001. Post Keynesian pricing theory reconfirmed? A critical review of Asking About Prices. *Journal of Post Keynesian Economics* : 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, J. 1983. *Oligopoly Theory*. Cambridge: Cambridge University Press.
- Friedman, M. 1953. The methodology of positive economics. In *Essays in Positive Economics*, Chicago: University of Chicago Press.
- Groenewegen, P. 1995. *A soaring eagle: Alfred Marshall 1842-1924*. Cheltenham: Edward Elgar.

- 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
- Keen, S. 2001. *Debunking Economics*, Sydney & London: Pluto Press & Zed Books.
- Keen, S. 2003. Standing on the toes of pygmies: why econophysics must be careful of the economic foundations on which it builds. *Physica A* 324: 108-116.
- Kuenne, R.E.1998. *Price and Nonprice Rivalry in Oligopoly*. London: Macmillan.
- 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*, Armonk: ME Sharpe.
- Lee, F.1998. *Post Keynesian Price Theory*. Cambridge: Cambridge University Press.
- Lipsey, R. G. & Chrystal, K. A. 1999. *Principles of Economics: 9e*. Oxford: Oxford University Press.
- Mankiw, N.G. 2001. *Principles of Microeconomics*. New York: Harcourt.
- Marks, R. 2000. Evolved Perception and the Validation of Simulation Models. In Barnett, W., Chiarella, C., Keen, S., Marks, R., Schnabl, H. (eds.), *Commerce, Complexity and Evolution*, New York: Cambridge University Press.
- Marshall, A. 1907 [1961]. *Principles of Economics*. London: Macmillan.
- Marshall, A. Correspondence with Professor A. L. Bowley. In Pigou, A.C., (ed.), (1925). *Memorials of Alfred Marshall*, London: Macmillan Publishing Company.
- Martin, S. 2001. *Advanced Industrial Economics*, Cambridge:Blackwell.
- Mas-Colell, A. (1983. Walrasian Equilibria as Limits of Noncooperative Equilibria. Part I: Mixed Strategies. *Journal of Economic Theory* 30: 153-70.
- Makowski, L. 1987. Are 'Rational Conjectures' Rational? *The Journal of Industrial Economics* 36: 35-47.
- Means, G.C. 1972. The administered-price thesis reconsidered. *American Economic Review* 62: 292-306.
- Moss, S. 1984. The history of the theory of the firm from Marshall to Robinson to Chamberlin: the source of positivism in economics. *Economica* 51: 307-318.
- Nichol, A. J.1934. A Re-Appraisal of Cournot's Theory of Duopoly Price. *The Journal of Political Economy*, 42: 80-105.
- Novshek, W. 1980. Cournot Equilibrium with Free Entry. *Review of Economic Studies* 47: 473-86.
- Novshek, W. 1985. Perfectly Competitive Markets as the Limits of Cournot Markets. *Journal of Economic Theory* 35: 72-82.

- Novshek, W. & Sonnenschein, H.1983. Walrasian Equilibria as Limits of Noncooperative Equilibria. Part II: Pure Strategies. *Journal of Economic Theory* 30: 171-87.
- 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.
- Rospit, P.G. 1993. The limits to deregulation of entry and expansion of the US gas pipeline industry. *Utilities Policy* October: 287-294.
- Rotwein, E.1959. On 'The Methodology of Positive Economics'. *The Quarterly Journal of Economics* 73: 554-575.
- Sato, K.1979. A note on capital and output aggregation in a general equilibrium model of production. *Econometrica* 47: 1559-1568.
- Schmalensee, R.1988. Industrial Economics: an overview. *The Economic Journal* 98: 643-681.
- Schumpeter, J. A.1936. *The Theory of Economic Development*. Cambridge MA: Harvard University Press.
- 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.
- Simon, J.L., Puig, C.M., & Aschoff, J. 1973. A duopoly simulation and richer theory: an end to Cournot. *Review of Economic Studies* 40: 353-366.
- 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.

¹ I would like to thank John Legge, (La Trobe University), Geoff Fishburn (University of New South Wales), Russell Standish (University of New South Wales), Mike Honeychurch (University of Melbourne), Robert Vienneau (ITT Industries Advanced Engineering Division), Trond Andresen (Norwegian Institute of Technology), Joanna Goard (University of Wollongong), Michael Kelly (Illinois Institute of Technology), Karol Gellert (University of Western Sydney), Raja Junankar (University of Western Sydney), Greg Ransom (UCLA Riverside) and members of the HAYEK-L internet discussion list for useful insights and comments on earlier drafts.

² This doesn't raise issues about "movements of the supply curve versis shifts along it" because that analysis

presumes the market is in equilibrium at the intersection of supply and demand. The question considered here is instead is “what happens to market price if the quantity supplied changes?” without any assumption about the current state of the market. Consider cobweb models of price-setting dynamics: the movement from one market quantity to another in those models involves no shift in the “supply curve”, but obviously involves changes in the output levels of individual firms.

³ This doesn't mean that the amount other firms vary their output by is always zero, or that the other firms will not in time respond to the impact that the change in the i^{th} firm's behavior has on the overall market; only that the amount of variation in direct and immediate response to a change in output by the i^{th} firm is zero.

⁴ This relation also causes problems when one wishes to compare the welfare effects of different industry structures, as I discuss below.

⁵ As Stigler argued, if all firms equate their own-output marginal revenues and marginal costs then the limit of $(P(Q)-MC(Q))$ is zero as the number of firms rises indefinitely, and this limit is approached from above so that $(P(Q)-MC(Q))$ is always non-negative.

⁶ I return to what cost functions can be if they are to be comparable later.

⁷ Why these conditions are restrictive is discussed in the next section.

⁸ Economies of scale in research and technology will also favor larger-scale operations. While these are ruled out in the standard comparison with identical marginal cost curves, they cannot be ruled out in the general case of differing marginal cost curves.

⁹ This will apply to any individual firm that changes its output; though of course the impact of its change will be felt by all firms in the industry.

¹⁰ The former could perhaps be undertaken using game theory, in which perfect competition is the limit of the Prisoners' Dilemma “defect” strategy as the number of firms rises indefinitely. However, it is well-known that the iterated Prisoners' Dilemma does not necessarily converge to the defection solution (Schmalensee 1988: 647). Secondly, there are serious dimensionality issues in extending the analysis of single period 2×2 player/strategy combinations to those with multiple periods, n players, and multiple feasible coalitions and strategies (Marks 2000). Leaving aside the issue of coalitions, the number of potential states scales as $\text{Strategies}^{\text{Players} \cdot \text{Rounds}}$. A 50 person 3 period game has over 10^{33} possible outcomes. It is difficult to sustain that competitive firms could be strategically analyzing this number of outcomes. Thirdly, no alternative route can circumvent the marginal cost aggregation problem, so that any analytic conclusions (such as the welfare-superiority of competition over monopoly) would remain tentative without empirical evidence on the cost functions of competitive and monopoly firms.

¹¹ If this is hard to swallow, read Chapter 3 of *Debunking Economics* where I explain Sraffa's impeccable logical

argument against the concept of diminishing marginal productivity.

¹² These include the need to cover fixed costs at a levels of output well within production capacity, the desire to finance investment and/or repay debt with retained earnings, the impact of the trade cycle, and the degree of competition (so that empirical research gives some grounds by which a more competitive industry can be preferred to a less competitive one)