Week 7 - Perfect Competition and Monopoly

Our aim here is to compare the industry-wide response to changes in demand and costs by a monopolized industry and by a perfectly competitive one. We distinguish between the short run, when the number if firms in the industry is fixed, and the long run, where new firms can enter or exit in the perfect competition case, and where a loss making monopolist can shut down in the monopoly case. We are particularly interested in the consequences of these equilibria for consumer welfare and producer profits.

We will first characterize the monopolist's pricing behaviour with respect to the concepts of elasticity and marginal revenue. Let the amount of good x demanded be q_x and the price of good x be p_x. Suppose that our market demand function is $q_x=D(p_x)$ (where the preferences – i.e. indifference curve map - and income of each consumer and the prices of all other goods are held fixed). Our inverse demand function will be $p_x=D^{-1}(q_x)$ Suppose that our firm's total cost function is given by $C(q_x)$ (where the production technology – i.e. isoquant map – and the prices of all inputs are held constant). Total revenue will be given by $R_x=q_x p_x=q_x D^{-1}(q_x)$ and marginal revenue will be given by:

$$MR_x= \partial R_x/\partial q_x= \partial/\partial x(q_x D^{-1}(q_x))$$
$$MR_x = q_x(\partial D^{-1}/\partial q_x)+ D^{-1}(q_x)$$
$$MR_x = q_x(1/(\partial D/\partial p_x))+ D^{-1}(q_x)$$

(*using the product rule and the fact that $\partial D/\partial p_x=1/(\partial D^{-1}/\partial q_x)$ because one function is the inverse of the other – inverse function rule*)

Average revenue will be given by $AR_x=R_x/q_x=p_x= D^{-1}(q_x)$. So, average revenue is simply the current market price. Remember that the calculus definition of price elasticity of demand (Varian's definition) is $\varepsilon_x=(\partial D/\partial p_x)(p_x/q_x)$. This means that (from the formulas for MR_x and AR_x above):

$$MR_x = p_x/\varepsilon_x + D^{-1}(q_x)$$
$$MR_x = AR_x\,(1/\varepsilon_x + 1)$$
$$\mathbf{MR_x = (1/\varepsilon_x + 1)\,AR_x}$$
$$MR_x = (1/\varepsilon_x + 1)\,p_x$$

This is another way of establishing the result that marginal revenue can only be positive (i.e. revenue can only increase along with the quantity sold) when $\varepsilon_x<-1$ (i.e. along the elastic part of the demand curve). When $\varepsilon_x>-1$ (but, remember, always less than 0 for a normal good on Varian's definition), the $(1/\varepsilon_x + 1)$ part is always negative and so revenue decreases as the quantity is increased (this is along the inelastic part of the demand curve). In the case of a linear demand curve, the price elasticity of demand is less than -1 in the top half of the demand curve and greater than -1 in the bottom half of the demand curve. By substituting in the formula for price elasticity of demand along a linear demand curve $\varepsilon_x=-b(p_x/q_x)$ we can find the formula for the marginal revenue curve along a linear demand curve:

$$MR_x = AR_x\,(1/\varepsilon_x + 1)$$
$$MR_x = p_x\,(1/(-b(p_x/q_x)) + 1)$$

$$MR_x = p_x - (1/b)q_x$$
(using the formula for a linear demand curve $q_x=b(a-p_x)$)
$$MR_x = 2p_x - a$$

From this formula we can see explicitly that marginal revenue will be positive when $p_x > a/2$ and negative when $p_x < a/2$. So, we now have enough information to sketch what the marginal revenue, demand curve and elasticity of demand curves will look like for a linear demand curve:
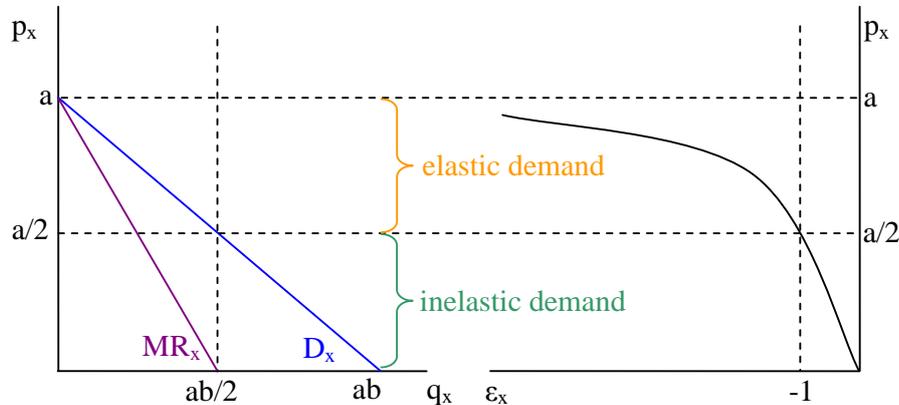
The profit function of a monopolist in the market for good x will be given by $\pi_x = R_x - C_x$ (i.e. profit is total revenue minus total cost). The derivative of the profit function with respect to q is the marginal increase in profit when quantity sold is increased. The monopolistic firm will wish to expand its output until $\partial \pi_x / \partial q_x = 0$:

$$\partial \pi_x / \partial q_x = \partial R_x / \partial q_x - \partial C_x / \partial q_x = 0$$
$$\partial R_x / \partial q_x = \partial C_x / \partial q_x$$
$$MR_x = MC_x$$

The marginal cost is always positive, so for this condition to be met, the marginal revenue must also be positive. This in turn implies that the elasticity must be less than -1; i.e. the monopolist will always choose to produce an output level on the elastic part of the demand curve. Intuitively, if the monopolist is on the inelastic part of the demand curve, they can increase their revenue by reducing the quantity supplied to the market. Since this also saves on production costs, it will never be rational for a profit maximizing firm to produce on the inelastic part of the demand curve.

The second order condition for profit maximization requires that:

$$\partial^2 \pi_x / \partial q_x^2 = \partial^2 R_x / \partial q_x^2 - \partial^2 C_x / \partial q_x^2 < 0$$
$$\partial MC_x / \partial q_x > \partial MR_x / \partial q_x$$

This condition tells us that marginal cost must be climbing faster than marginal revenue at the point where they cross if it is to be a profit maximum. In the case of the perfectly competitive firm introduced below, since MR_x simply becomes P_x (because the firm is a price-taker) then $\partial MR_x / \partial q_x$ becomes 0 and we simply have the condition that MC must be increasing at the crossing point.

In the following analysis, we will assume that the monopolist, or each individual firm in the case of the perfectly competitive industry, has a production technology which exhibits decreasing returns to scale (i.e. increasing marginal cost) at all output levels and has a fixed cost F, so that total cost $C_x = V_x + F$ where V_x is the variable cost so that $\partial V_x / \partial q_x = MC_x$ (because the derivative of F with respect to q_x is 0). This assumption ensures that there is a point of minimum average cost, where all economies of scale have been exhausted and diseconomies

of scale are beginning to set in. We also assume a linear demand curve so that we can be precise about the shape of the marginal revenue curve.

   A profit maximizing firm will always maximize its profits by setting marginal revenue equal to marginal cost. The marginal revenue curve for a monopolist has been discussed above. The area above the marginal cost curve and below the market price is in fact simply the producers' surplus; it is the difference between the price the firm gets for each additional unit and the minimum amount (assuming it has paid its fixed cost F) that the firm is willing to supply each additional unit for. To find the firm's profits, we must take the producers surplus and subtract the fixed cost F.

   The diagram below illustrates the case of a monopolistic firm. Since the monopolist faces the entire market demand curve and can freely choose any point on it (this is not the case with a perfectly competitive firm, as we shall shortly see) they will maximize their profits by choosing quantity q_x' where marginal revenue equals marginal cost. The producers' surplus is the horizontally hashed roughly trapezoid area. The monopolist's profit is equal to the producers' surplus minus the fixed cost. It can be shown on the graph as the vertically hashed rectangle because since p_x' is the average revenue, then the height of the rectangle is average revenue minus average cost and the width of the rectangle is the amount produced, so the area of the rectangle is total revenue minus total cost.

   The area of the rectangle is sometimes referred to as the monopolist's **economic profit**. This is because we assume that the marginal cost curve includes all of the factors of production, including managerial and entrepreneurial ability. So, the idea is that there is a normal "wage" for entrepreneurs and managers, which could be quite high if these skills are in short supply. The **economic profit** is the extra income (received by the owner of the firm, which could be the entrepreneur, or the shareholders, which will probably include among their number the managers) which derives from the monopolist's **monopoly power** or **market power**, which refers to the monopolist's ability to set the price level to whatever it chooses without fear of other firms competing with it by setting prices lower. The price and output quantity which would maximize total surplus (i.e. would be socially optimal) is (q_x^*, p_x^*). This market power, as well as leading to positive economic profits for the monopolist, leads to a loss in consumer welfare (since the monopolist's price p_x' is higher than the socially optimal output quantity p_x^* the consumers' surplus must be lower) and a **deadweight loss**, due to the distortion of the market below the socially optimal quantity This is shown as the diagonally hashed area in the graph below.
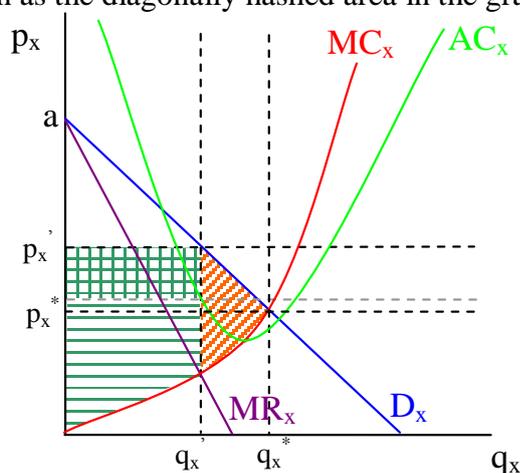
The diagram below illustrates the effect on the monopolist's economic profits and on consumers' surplus when the demand curve shifts outwards. The outwards shift in the demand curve shifts the marginal revenue curve as well (so that it still touches the x axis half way between the origin and the point where the demand curve touches the x-axis). There is therefore an increase in the quantity sold but also the market price. This means that the monopolist is able to translate some of the increased demand into a higher price for its goods (as we shall shortly see, this cannot occur in the long run equilibrium of a perfectly competitive industry). The monopolist's profits clearly increase from the vertically hashed rectangle to the vertically and horizontally hashed rectangle (including the small double hashed part). The effect on consumers' surplus is ambiguous because although the consumer's gain the north-east-south-west diagonally hashed part on the right, they lose the vertically and horizontally hashed part, which is transferred to the monopolist's economic profits. The size of these areas would depend on the shape of the demand curve, so we cannot draw any firm conclusions just from this diagram. There is also an ambiguous effect on the size of the deadweight loss.
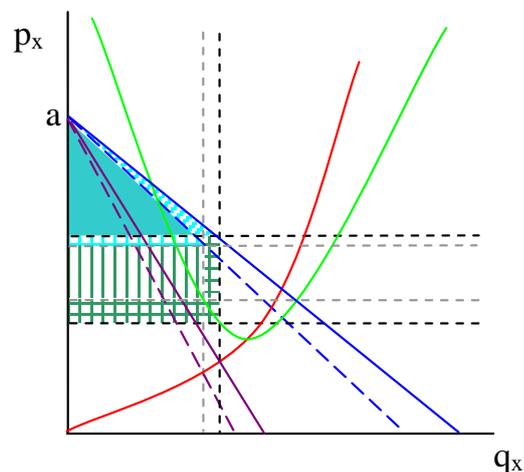
The standard model of perfect competition firm differs from that of a monopolist in the following ways:-

I. There are a large number of firms. This means that the output of each firm is tiny relative to the output of the industry as a whole. This is very important as a justification of the simplifying assumptions made. For simplicity, we assume that all firms have identical production functions (and therefore cost curves), and so each make up an equal part of total industry output, due to the symmetry of the model.

II. There is a distinction between the profit maximization decision of each individual firm and the market equilibrium in the industry as a whole. Each firm is modelled as having a perfectly elastic demand curve at the current industry market price. The idea is that if a perfectly competitive firm tries to charge a price above the current industry market price its demand immediately goes to 0 because consumers will be able to buy from other firms at the industry market price, and so will all abandon the firm which raised its price. Similarly, if the individual firm charges a price lower than the current industry market price, customers will flood

to it from all the other firms, and because the individual firm is so small relative to industry output, to all intents and purposes demand is perfectly elastic.

III. There is a distinction between the short run and long run responses to changes in demand or supply conditions. In the short run, individual firms alter their behaviour, and this is reflected in a new equilibrium for the industry as a whole. The number of firms, however, is considered fixed in the short run. In the long run, on the other hand, we assume that new firms can enter the market. They will do this whenever the firms currently in the market are making positive economic profits. This is because the positive economic profits enable firms to make a higher rate of profit than the standard requirements built into the marginal cost function (i.e. if entrepreneurs move into this industry, they will get a higher return on their skills than the average elsewhere in the economy). By the same token, firms will leave an industry where they are making negative economic profits. This means that the industry can only be in long run equilibrium where all firms are making 0 economic profits. Since the demand curve faced by each individual firm is perfectly elastic, it must always be the case that $AR_x = p_x = MC_x$. Since 0 economic profits can only occur when $AR_x = AC_x$ this in turn implies that at the long run industry equilibrium $p_x = MC_x = AC_x$. So, the long run equilibrium price is where the marginal cost and average cost curves cross, i.e. the point of minimum average cost. Intuitively, new firms will enter the industry, or existing firms leave the industry, until each firm is producing at its minimum average cost. This is the only situation where there are no profitable opportunities for new firms to enter the market and undercut the prices of the existing firms.

The diagram below illustrates the effect of an outward shift in the demand curve for a perfectly competitive industry. Suppose that the industry is initially in long run equilibrium so that each firm is producing q_x^* units and there are N firms so that overall industry supply is Nq_x^* units, which is the amount demanded by consumers along demand curve D_x at price p_x^*. Each firm is initially making 0 economic profits. Suppose the demand curve then shifts to $D_x^{'}$. Since the number of firms is fixed in the short run, the short run supply curve SS_x must be upward sloping because each firm experiences decreasing returns to scale. It must also go through the initial long run equilibrium because the number of firms is fixed at the level compatible with long run equilibrium at the intersection of the LS curve with the initial demand curve D_x. So, the new short run industry price will be $p_x^{'}$. Each individual firm will have increased output to $q_x^{'}$ and will be making economic profits in the short run equal to the shaded rectangle. In the long run, the number of firms will increase to N' and each firm will again be producing q_x^* units so that there is once again a long run industry equilibrium with 0 economic profits for each firm.
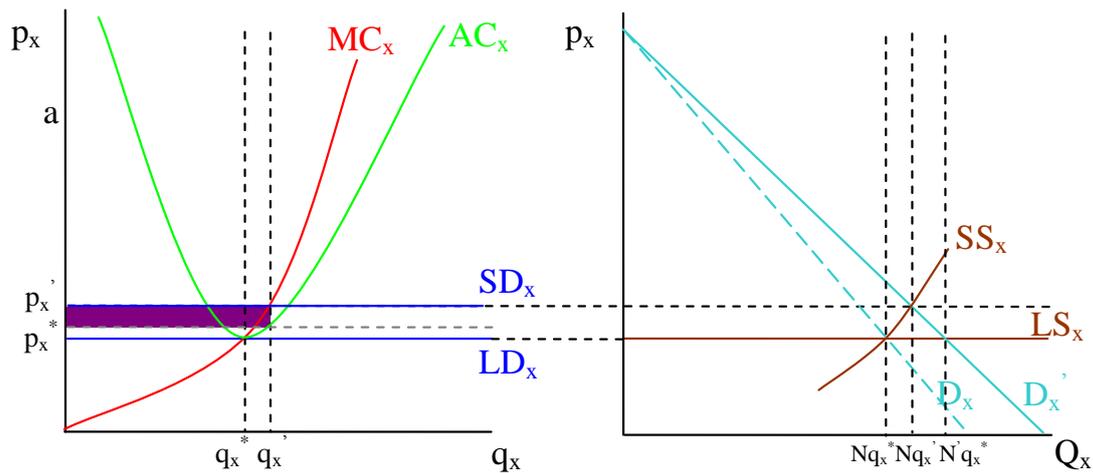
Once we have reached the new long run equilibrium, consumers' surplus will have unambiguously increased, and industry economic profit will remain unchanged at 0.

An increase in workers' wages is represented by an upward shift in the marginal and average cost curves. This will clearly lead to an increase in price and a reduction in output quantity in the case of a monopolist because the intersection of MC with MR must now be at a lower quantity. Consumers' surplus therefore clearly decreases. The effect of the monopolist's profit will also be unambiguously negative because the total cost of production at any quantity will increase, and so the monopolist's profits will be reduced whatever output they choose. In the diagram below, the monopolist's economic profits go from the diagonally hashed area to the smaller dark solid rectangle.
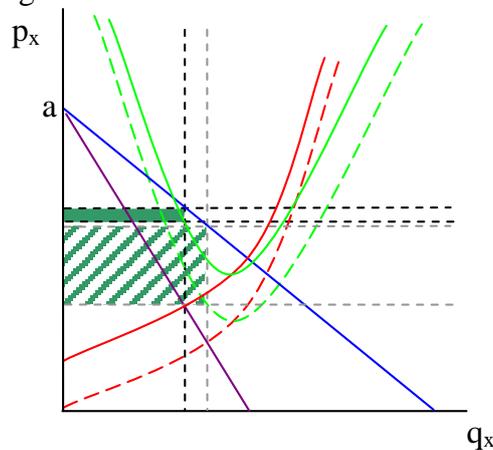
In the case of a perfectly competitive industry, the increase in workers' wages shifts the long run industry supply curve upwards because the minimum average cost of production increases from p_x^1 to p_x^2. The short run supply curve is upwards sloping because each firm has decreasing returns to scale and goes through the point on the new long run supply curve which corresponds to the initial N firms producing q_x^* output each and selling at price p_x^2. Since the number of firms is fixed in the short run, the short run equilibrium requires that each firm reduce its output from q_x^* to q_x' and that the price increase to p_x' (resulting in reduced consumer surplus). Each individual firm is now making negative economic profits equal to the area of the dark rectangle. The long run equilibrium involves firms leaving the industry until N' firms are producing q_x^* units of output each and selling for price p_x^2. Consumers' surplus

has clearly been reduced. Overall industry economic profits have returned to their long run equilibrium level of 0.
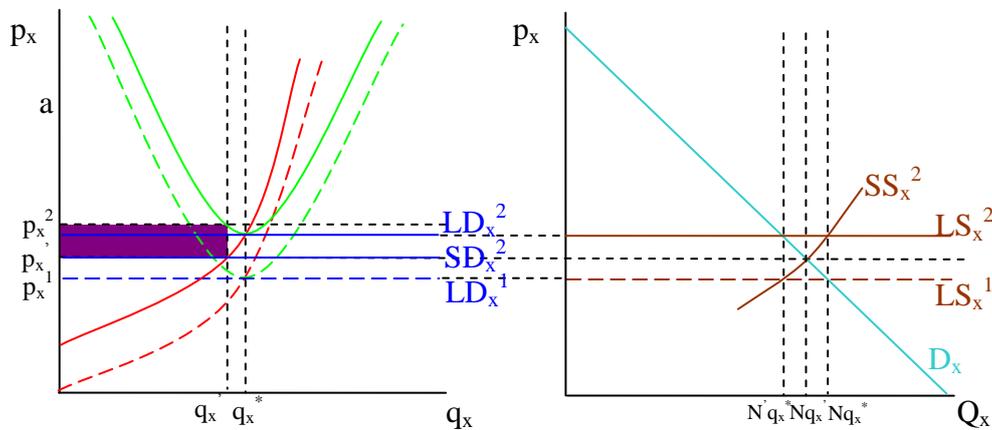
If the proportional tax on profits falls only on economic profits, this will have no effect on the behaviour of either a monopolistic or a competitive industry in turns of the above analysis. Regardless of whether the firm is a monopolist or a competitive one, its after-tax economic profits are given by $\pi_x^T=(1-t)\pi_x=(1-t)(R_x-C_x)$ where t is the tax rate and π_x is pre-tax economic profits. A profit-maximizing form sets:

$$\partial\pi_x^T/\partial q_x =(1-t)(\partial R_x/\partial q_x - \partial C_x/\partial q_x) = 0$$
$$(1-t)\ \partial R_x/\partial q_x = (1-t)\ \partial C_x/\partial q_x$$
$$MR_x = MC_x$$

So, the proportional tax on economic profits has no effect on the profit maximization condition of marginal revenue equals marginal cost. The proportional tax will reduce the firm's economic profits by a certain proportion but will have no effect on its price or output setting behaviour.

In the case of the perfectly competitive industry, the long run equilibrium requires 0 economic profits, so a proportional tax on economic profits will clearly have no effect on the long run equilibrium.

Supposing as an alternative that the proportional tax falls not just on economic profits but on the "wage" paid to entrepreneurs or the normal rate of return on the capital owned by the firm. (This would likely occur in the real world because accounting profits include entrepreneurial wages and the entire earning from the firm's capital stock.) This would have a similar effect on marginal, average and total costs as the increase in workers' wages, and so the same diagrams would apply.

We saw earlier that the existence of monopoly power creates a deadweight loss of total social welfare because the monopolist sets $MR_x=MC_x$ instead of $p_x=MC_x$ which would be required if the marginal utility (strictly speaking, assuming that we have quasi-linear utility) of the consumer from consuming the final unit is to be equated with the marginal cost of producing the final unit. Clearly, on the analysis so far, monopoly is a bad thing for society, not only because of the total loss of welfare, but because the economic profits generated by monopoly power will tend to be transferred from consumers to a small group (e.g. shareholders) and thus also increase the inequality of the utility generated by the economy as well as reducing the total amount of it.

There is an area of controversy in economics about the true extent of monopoly power, how damaging it really is, and how much the government can or

should intervene to prevent it. We will shortly see alterations to the model which may lead to the conclusion that monopoly is not as bad as the simple model would suggest. Firstly, however, we should consider the reasons why monopoly power may exist:

1) **Technological** – If we look at the diagram for a firm with economies of scale at high output levels below, we can see that if there were to be two firms in the market instead of 1, there would simply not be enough market demand for both firms to produce at their minimum average cost. If both firms tried to produce $(1/2)q_x^*$ each so that the overall quantity produced is the optimal amount for the single monopolist to produce, the average cost for each firm would be above the price, and so both would make a loss. This is because both firms are now paying the fixed cost F, whereas only one was before. It would also not be optimal from society's point of view for both firms to pay F. The best thing from the point of view of maximizing total surplus would be for a single firm to produce, and for that firm to be forced to produce q_x^* so that price equals marginal cost. This is roughly what happens with regulated private monopolies such as water companies in the UK.
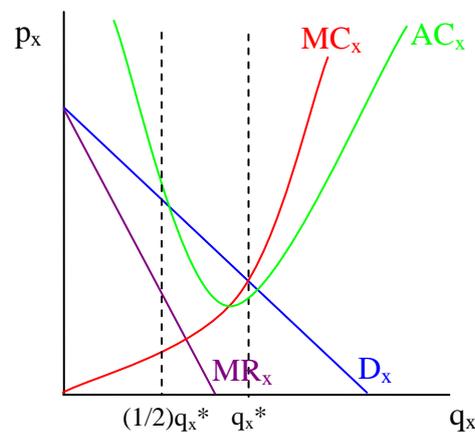

2) **Barriers to entry** - The monopolist may be the only firm able to produce its particular product, either because it has legal protection such as a patent or government contract, or, for example, because it has a closely guarded secret product design. Firm's may also use strategic ploys such as building up excess capacity so that if any other firm tries to challenge them they will flood the market and thus drive the challenger out of business.

3) **Product differentiation** – There may be a big difference between the products produced by different companies in the same industry (e.g. car firms). This will mean that whilst the different firms' products are substitutes, they are not perfect substitutes (which is the assumption required for the perfect competition model) and so each firm has some monopoly power.

If a monopolist were able to discriminate perfectly between consumers, it would be able to charge a highest price each consumer would be willing to pay for each marginal unit. This means that the monopolist's producers' surplus would be the entire area between the demand curve and the marginal cost curve. The monopolist would be able to gain the whole of what would have been the consumers' surplus because as you go along the demand curve it charges each marginal consumer exactly

what they would have been willing to pay. This also means that by expanding production beyond the optimal non-price-discriminating quantity, the monopolist is able to gain what would have been the deadweight loss in the non-price-discriminating case. So, a monopolist able to perfectly price discriminate produces at the socially optimal output level, and there is therefore no deadweight loss, i.e. total surplus is maximized. Consumers are, however, made worse off as they lose the turquoise diagonally hashed triangle which was their consumers' surplus in the non-price discriminating case. This suggests that although perfect price discrimination is good from the point of view of maximizing total welfare, it will cause even more inequality than non-price-discriminating monopoly power because the small groups who own the monopolistic firm will be able to take all of the welfare which would have gone to the consumers if we had had a single market price equal to marginal cost (as suggested the government should do in the question above). So even if perfect price discrimination were possible, government regulation of monopoly pricing would probably still be preferable. It is very unlikely that perfect price discrimination could be practical in reality anyway because it would involve the monopolist knowing exactly how much each consumer would be willing to pay. Other forms of price discrimination are more likely to be achievable in practice, but with these a deadweight loss from monopoly power will remain.
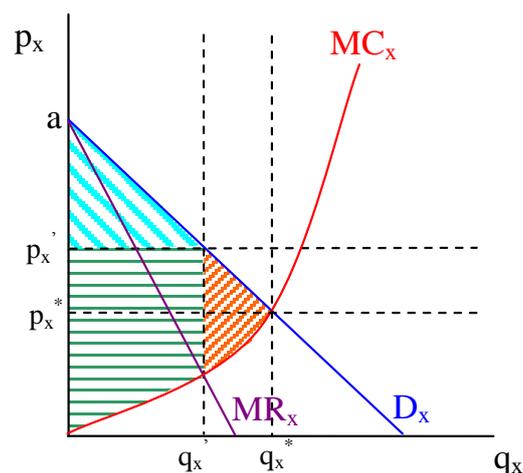
We have so far looked at the **allocative inefficiency** caused by monopoly power, which refers to the socially suboptimal amount of output produced by a monopolist. Another common result of monopoly power is **productive inefficiency**. This refers to a failure of a monopolist to minimize cost of production. This is similar in effect to the analysis earlier on where we saw that higher prices for inputs will reduce the economic profit of a monopolist. The problem arises when the group receiving wages as part of the marginal cost curve is the same as the group which receives the economic profits. It may then be desirable for that group to push up the marginal cost curve because it will give them extra income in excess of the loss of economic profit. This cannot occur in the long run equilibrium in a competitive industry because any firm which does not produce at the minimum average cost available with the existing technology will be undercut by new competitors.

It is generally accepted that the government should try to remove as many barriers to entry as possible; it should not grant any special licenses unless absolutely necessary (e.g. for consumer safety, as with the medical industry). However, it is

sometimes argued that monopoly in high tech sectors is a good thing because it gives firms a greater incentive to invest in research and development, because they will reap the benefits as a monopolist. However, there are better ways for the government to encourage research and development than by tolerating or deliberately fostering monopoly.

The exception to this argument is that of patents for new ideas. Although R&D can be encouraged by directly targeted government subsidies, the more specific a new idea, production technique or product is, the more it is likely to be necessary to allow the developer to enjoy the exclusive use of it for a certain finite period so that they can recoup their initial expenses and be given an incentive to direct the investment in the first place. The government is not likely to succeed in the business of picking which specific ideas to fund because it has less information than the sum of all agents in the market. However, there will always be a finite time period for an optimal patent because sooner or later the rents accruing to the firm with the patent are sufficient to provide the required incentive and the static welfare losses from monopoly power become a pure deadweight loss even from this **dynamic efficiency** perspective.

The technological causes of monopoly power (high fixed costs) often mean that it would be either undesirable or impossible for the government to deliberately put more firms in the industry than is socially optimal. We have already seen that there are other ways to alleviate the negative effects of monopoly power. These include regulating the monopolist's price and ensuring that if there is a small number of firms that these are prevented from colluding to keep prices high (this is the role of competition policy). It may be possible through a well designed competition policy to make even a small number of firms behave competitively, as would be predicted by the Bertrand model of duopoly with homogenous goods and identical costs.