Monopoly

Pablo Mollá Chárlez

February 27, 2025

Contents

1 Profit Maximization in a Monopoly **2**

1.1 Marginal Revenue = Marginal Cost . . . 2

1.2 Marginal Revenue and Elasticity . . . 2

1.3 Why a Monopolist Avoids the Inelastic Region . . . 3

1.4 Example: Linear Demand Curve and Monopoly . . . 3

1.5 Markup Pricing . . . 4

1.6 Example . . . 4

2 Inefficiency of Monopoly: Pareto Inefficient **5**

3 Natural Monopoly **7**

4 What causes Monopolies = State **8**

4.1 Minimum Efficient Scale . . . 8

4.2 Government's Role in Market Size . . . 8

4.3 Libertarian Perspective . . . 8

4.4 Example . . . 9

1 Profit Maximization in a Monopoly

Let $p(y)$ be the demand function, giving the price at which y units of output can be sold. Let $c(y)$ be the cost function, and let $r(y) = p(y)\,y$ be the revenue function. A monopolist chooses y to solve:

$$\max_{y} \ \pi(y) \ = \ r(y) \ - \ c(y).$$

1.1 Marginal Revenue = Marginal Cost

To find the profit-maximizing output, we set marginal revenue (MR) equal to marginal cost (MC):

$$\frac{\partial}{\partial y}\big[r(y)\big] \ = \ \frac{\partial}{\partial y}\big[c(y)\big] \ \longleftrightarrow \ MR(y) \ = \ MC(y).$$

In a competitive market and as we have seen in chapter 18, marginal revenue for each firm equals the market price p ($MR(y) = p$), so the condition reduces to $p = MC$. For a monopolist, however, marginal revenue is more subtle because increasing output affects both:

- **Quantity Effect:** Selling an additional Δy units at price p contributes roughly $p\,\Delta y$ to revenue.
- **Price Effect:** Raising output Δy lowers the price by Δp, and all units (not just the new ones) are sold at this lower price. This contributes $y\,\Delta p$ to the change in revenue (and note Δp is negative if price falls).

Hence, if the monopolist changes output by Δy,

$$\Delta r \ = \ p\,\Delta y \ + \ y\,\Delta p,$$

and dividing by Δy yields the marginal revenue:

$$MR(y) \ = \ \frac{\Delta r}{\Delta y} \ = \ p \ + \ y\,\frac{\Delta p}{\Delta y}.$$

For a competitive firm, the market price is constant with respect to that firm's output—effectively $\frac{\Delta p}{\Delta y} = 0$. Thus the firm's marginal revenue equals p. So in competition:

$$MR(y) \ = \ p, \quad \text{and the optimal condition } p = MC.$$

1.2 Marginal Revenue and Elasticity

Recall the demand function $p(y)$. Let $\varepsilon(y)$ be the price elasticity of demand at output y, defined (in absolute value) as

$$\varepsilon(y) \ = \ -\frac{\partial y}{\partial p}\,\frac{p}{y} \quad \text{(or equivalently, } \frac{\partial p}{\partial y}\,\frac{y}{p} \text{ with sign considerations).}$$

A standard result (from Chapter **Market Demand**) is that for a monopolist:

$$MR(y) \ = \ p(y)\Big[\,1 + \frac{1}{\varepsilon(y)}\Big].$$

Setting $MR(y) = MC(y)$ implies

$$p(y)\Big[\,1 + \frac{1}{\varepsilon(y)}\Big] \ = \ MC(y).$$

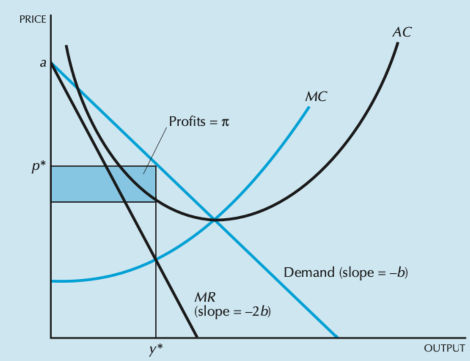
In contrast to a monopolist scenario, in a competitive environment, the firm faces (locally) an infinitely elastic demand curve due to the fact a slight 1% increase in price leads to more than a 1% decrease in quantity demanded as the other competitive firms would take the share of the market, i.e., $|\varepsilon| \to \infty$. Hence $1/\varepsilon \to 0$, so the condition collapses to $p = MC$.

1.3 Why a Monopolist Avoids the Inelastic Region

Suppose the demand curve is inelastic at some output y, meaning $|\varepsilon(y)| < 1 \longleftrightarrow \frac{1}{|\varepsilon(y)|} > 1$. Then $1 + \frac{1}{\varepsilon(y)} = 1 - \frac{1}{|\varepsilon(y)|}$ becomes a number larger than 1 in magnitude but negative in sign (since $\frac{1}{\varepsilon(y)} < -1$ for inelastic demand). In fact, one can show $MR(y)$ is negative if $|\varepsilon(y)| < 1$. A negative marginal revenue cannot equal a positive marginal cost, so such a point cannot be a profit-maximizing choice. In standard microeconomic models, marginal cost is assumed to be nonnegative (and typically strictly positive) for all positive output. The usual idea is that **producing more output requires at least some extra resources, so cost should not decrease as quantity increases**.

Intuitively, think of it this way: if the monopolist is currently producing in a region where demand is inelastic ($|\varepsilon| < 1$), reducing output actually raises total revenue (because the price increase on each unit more than makes up for selling fewer units). Meanwhile, the monopolist's total cost goes down because it's producing fewer units. Both effects—higher revenue and lower cost—push profit up. So if the monopolist ever finds itself in the inelastic portion of the demand curve, it can keep cutting output and keep raising its profit. That means no point in that inelastic region can be a final, profit-maximizing solution. Ultimately, the monopolist will move to (or beyond) the point where demand becomes elastic or unit-elastic, and only there can it reach a stable maximum of profit.

1.4 Example: Linear Demand Curve and Monopoly



Monopoly with a linear demand curve. The monopolist's profit-maximizing output occurs where marginal revenue equals marginal cost.

Figure 1: Monopoly Profit

The "profit box" in the figure 1 represents π = (price per unit) – (average cost per unit), all multiplied by the quantity sold, y^*. In symbols:

$$\pi \;=\; p^* \times y^* - \underbrace{AC\big(y^*\big) \times y^*}_{=C(y^*)}.$$

Geometrically, the height of that rectangle is the difference between the monopoly price p^* and the average cost $AC(y^*)$, and the width is y^*. Multiplying those gives you total profit: total revenue $\big(p^* \cdot y^*\big)$ minus total cost $\big(AC(y^*) \cdot y^*\big)$. That's why the shaded region is the firm's profit in the diagram 1. Just as a small reminder, if we consider a linear demand curve as $p(y) = a - by$:

- The firm's revenue is $r(y) = p(y)\,y \;=\; (a - b\,y)\,y \;=\; a\,y \;-\; b\,y^2$.
- Taking the derivative of $r(y)$ with respect to y gives the marginal revenue (MR) defined as $MR(y) = \frac{d}{dy}[r(y)] \;=\; a - 2\,b\,y$.
- If the cost function is $c(y)$, then the marginal cost (MC) is $MC(y) = \frac{d}{dy}[c(y)]$.
- The average cost (AC) is $AC(y) = \frac{c(y)}{y}$.

1.5 Markup Pricing

We can use the elasticity formula seen previously for the monopolist to express its optimal pricing policy which is known as markup pricing is interesting because it shows exactly how a monopolist's market power and the elasticity of demand jointly determine the price. The formula:

$$p^* = \frac{MC(y^*)}{1 - \frac{1}{|\varepsilon(y^*)|}}$$

reveals that the monopolist charges a markup above marginal cost, and that this markup depends directly on the elasticity of demand. If demand is less elastic (smaller $|\varepsilon|$), the markup is larger, and the firm can charge a price further above cost; if demand is more elastic (larger $|\varepsilon|$), the markup is smaller. This makes it clear how the firm's pricing power is fundamentally constrained by consumers' sensitivity to price changes.

1.6 Example

Let's consider a numerical example using a linear demand and constant marginal cost to illustrate why a monopolist sets $MR = MC$ and how its chosen price ends up above marginal cost. The setup is the following one:

- **Demand Curve:** $p(y) = 100 - y$. This means if the monopolist produces y units, the market price is $100 - y$.
- **Cost Function:** $c(y) = 20\,y$. This implies a constant marginal cost (MC) of 20 for every unit.
- **Revenue:** $r(y) = p(y)\,y = (100 - y)\,y = 100y - y^2$.
- **Marginal Revenue (MR):** Differentiate $r(y)$ with respect to y:

$$MR(y) = \frac{\partial}{\partial y}[100y - y^2] = 100 - 2y.$$

- **Marginal Cost (MC):** Since $c(y) = 20y$, we have

$$MC = \frac{\partial}{\partial y}[20y] = 20.$$

The monopolist chooses y so that $MR(y) = MC$, this is the optimal output:

$$100 - 2y = 20 \;\longleftrightarrow\; 2y = 80 \;\longleftrightarrow\; y^* = 40.$$

So the profit-maximizing output is 40 units. Then, the optimal price at y^* is:

$$p(40) = 100 - 40 = 60.$$

And the marginal cost $MC = 20$. Hence the monopolist's price is \$60, while marginal cost is \$20. Notice that $p^* = 60$ is well above the MC of 20. This price–cost gap arises because, for a monopolist, marginal revenue is always below the price (here, $MR(40) = 20$), so the firm equates MR to MC at a point where $p > MC$.

Why $MR = MC$ instead of $p = MC$?

In perfect competition, each firm faces a constant market price, so marginal revenue = price = MC at equilibrium. A monopolist must lower its price on all units to sell one more unit, so the extra revenue from that additional unit is less than the price. Consequently, $MR < p$. To maximize profit, the monopolist sets $MR = MC$. Because $MR < p$, we end up with $MC = MR < p$. That's why the chosen price (here, 60) is above marginal cost (20).

If the monopolist tried to set $p = MC$, it would mean $p = 20$. Plugging that into the demand curve gives $y = 80$ units, but at $y = 80$, the marginal revenue would be $100 - 2(80) = -60$. A **negative marginal revenue** means that producing an additional unit lowers total revenue. The drop in price on all existing units outweighs the revenue gained from selling the extra unit, so revenue actually decreases when output increases. In this situation, $MR = -60$ means that if the monopolist produces one more unit beyond $y = 80$, total revenue drops by \$60. In other words, the loss in revenue from having to lower the price on all 80 existing units (plus the new one) outweighs the gain from selling that single extra unit.

Profit at y=80 and p=20:

Total Revenue (TR) $= p \times y = 20 \times 80 = 1600$,

Total Cost (TC) $= 20 \times 80 = 1600$,

$\pi = \text{TR} - \text{TC} = 0$.

Producing that much would reduce profits (to zero) because the firm would be selling many units at a low price. Instead, it's more profitable to produce fewer units (40) and sell each at a higher price (60). Producing 40 units is the monopolist's profit-maximizing output: at that quantity, its marginal revenue equals its marginal cost (both equal 20 in our example). In practical terms, each additional unit up to the 40th adds at least as much to revenue as it adds to cost. Once you go beyond 40, the marginal revenue goes below marginal cost, so adding more units would reduce profit rather than increase it.

Profit at y=40 and p=60:

Total Revenue (TR) $= 60 \times 40 = 2400$,

Total Cost (TC) $= 20 \times 40 = 800$,

$\pi = \text{TR} - \text{TC} = 2400 - 800 = 1600$.

Thus, the monopolist makes a profit of \$1600 by producing 40 units at a price of \$60. If the monopolist produces less than 40, let's say $y = 20$, then at those lower quantities marginal ($MR(20) = 60$) revenue exceeds marginal cost ($MC(20) = 20$), meaning each extra unit still adds more to revenue than it adds to cost.

Profit at y=20 and p=80:

Total Revenue (TR) $= 20 \times 80 = 1600$,

Total Cost (TC) $= 20 \times 20 = 400$,

$\pi = \text{TR} - \text{TC} = 1600 - 400 = 1200$.

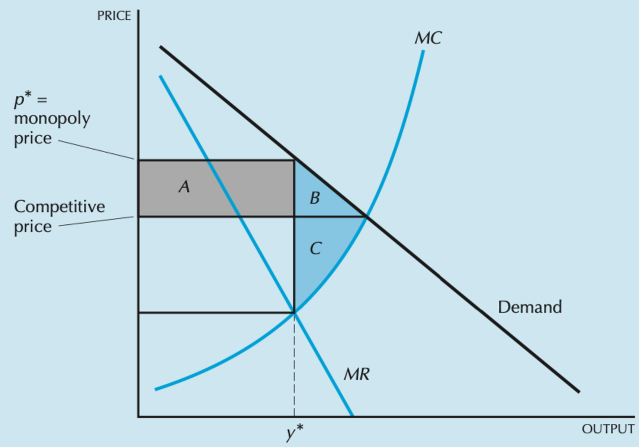
Thus, the monopolist makes a profit of \$1200 by producing 20 units at a price of \$80. Failing to produce those profitable units (because the monopolist could produce more) leaves money on the table, and as the monopolist produces fewer goods, total revenue decreases faster than costs fall, thereby reducing overall profit.

2 Inefficiency of Monopoly: Pareto Inefficient

In general the price will be higher and the output lower if a firm behaves monopolistically rather than competitively. For this reason, consumers will typically be worse off in an industry organized as a monopoly than in one organized competitively.

A monopolist chooses an output y_m where marginal revenue = marginal cost but price p_m is above marginal cost MC. This leads to lower output and higher price compared to the competitive outcome (p_c, y_c). From a Pareto efficiency standpoint, the monopoly outcome is inefficient because there remain mutually beneficial trades—units of the good—between y_m and y_c that do not occur under monopoly.

- Pareto Inefficiency of Monopoly: If the monopolist is producing y_m units, and for each extra unit between y_m and y_c the willingness to pay $p(y)$ exceeds the marginal cost $MC(y)$, there is a missed opportunity: a consumer would gladly pay more than it costs to produce that additional unit. By not producing those extra units, the monopolist leaves "money on the table" in the sense that both the consumer (who would pay above cost) and the producer (who would sell above cost) could be better off without harming anyone else. This is the hallmark of a deadweight loss and violates Pareto efficiency.



Deadweight loss of monopoly. The deadweight loss due to the monopoly is given by the area $B + C$.

Figure 2: Monopoly Deadweight Loss

- **Contrast with Perfect Competition:** In competitive markets, each firm produces until price p equals marginal cost MC. Here, the willingness to pay for the last unit (which is p) just matches the cost of producing that last unit. **This maximizes total surplus** (the sum of consumer surplus and producer surplus) **and leaves no unexploited opportunities** where a consumer would pay more than cost. Hence, the competitive equilibrium is Pareto efficient under standard assumptions (no externalities, well-defined property rights, etc.).

- **Why the Monopolist Produces Less?** The monopolist also considers the price effect on all infra-marginal units (the units already being sold) when it increases output. Selling an extra unit requires lowering the price for all units (uniforming pricing), which can reduce overall revenue from the existing sales. This extra factor is irrelevant to pure efficiency but very relevant to the monopolist's profit calculation. As a result, the monopolist cuts back on output to avoid lowering the price on all the previously sold units.

Hence, a monopoly typically sets $y_m < y_c$ and $p_m > p_c$. The lower output level creates a deadweight loss: there are consumers who value the product above its marginal cost but can't purchase it at the monopoly's higher price. By contrast, in a free-market competitive setting, firms produce until $p = MC$, no beneficial trades are left unmade, and the outcome is Pareto efficient.

$\boxed{\textbf{Important:}}$

- Competitive Firm: It takes the market price as given—it cannot affect that price by changing its own output. So its profit-maximizing decision is:
  - Observe the Market Price p.
  - Choose Quantity q such that marginal cost $MC(q) = p$ and produces more if $p > MC$.
  - Accept that the total revenue is $p \times q$ at that given price.
  - No "price effect" on previously sold units, since the firm's output is too small to affect the market price.

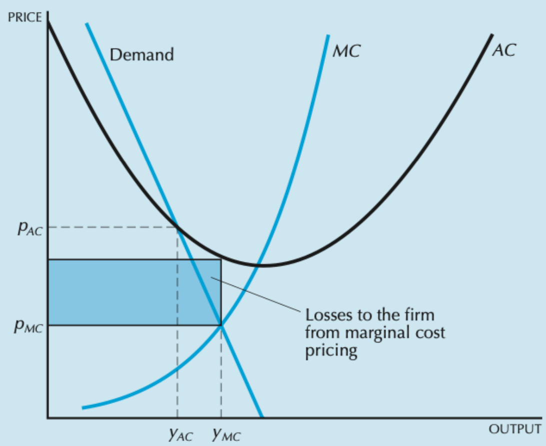
- Monopolist: It faces the entire market demand curve and does affect price by changing output:
  - Choose Quantity y.

– Set Price $p(y)$ as determined by the demand curve.
– Collect total revenue $p(y) \times y$, knowing it must lower price on all units if it wants to sell more.

It is possible for a competitive firm to earn a profit even if $p = MC$. Remember that profit $(p \times y - AC(y) \times y)$ depends on average cost (AC - the total cost per unit), not just marginal cost (MC - additional cost of producing one more unit). A firm can make a positive profit if price exceeds average cost, even if $p = MC$.

3 Natural Monopoly

A natural monopoly arises when a firm's technology has very large fixed costs (e.g., building an infrastructure network) but very small marginal costs (e.g., the cost of serving an extra customer). In such cases, a single firm can often supply the market's demand at lower total cost than multiple competing firms. Traditional examples include water, electricity, gas, or telephone networks: once you've laid pipes or cables, providing each additional unit costs little, but you still have enormous fixed expenses to recover.



A natural monopoly. If a natural monopolist operates where price equals marginal cost, then it will produce an efficient level of output, y_{MC}, but it will be unable to cover its costs. If it is required to produce an output where price equals average cost, y_{AC}, then it will cover its costs, but will produce too little output relative to the efficient amount.

Figure 3: Natural Monopoly

In Figure 3, this large fixed-cost structure means that marginal cost (MC) lies below the firm's average cost (AC) for much of the relevant range of output. The truly efficient (Pareto-efficient) amount of output would occur where price (demand) = MC (labelled y_{MC} in the diagram). But at that point, the average cost is still above price, so the firm would lose money—eventually going out of business **unless subsidized**. Traditional regulation or government ownership tries to fix this by either:

- **Setting price = AC**, ensuring the firm breaks even but still produces less than the efficient level ($y_{AC} < y_{MC}$).
- **Setting price = MC** and subsidizing the firm's losses so it can cover its large fixed costs.

From a libertarian and free-market perspective, however, there's no need for state intervention or protection. If a firm truly has the lowest cost structure (due to economies of scale and scope), it will naturally become the sole provider—yet it must still offer attractive prices to keep potential competitors out. In a genuinely free market (with no government licenses or legal barriers restricting new entrants), if the "natural monopolist" ever tried to exploit its position by pushing prices too high, other entrepreneurs could set up parallel infrastructure or alternative technologies. The threat of this new competition forces the incumbent to price near its true costs, ultimately benefiting consumers with better services and lower prices. Under these conditions:

- No forced entry barrier: If profit margins are high, rivals can invest in their own network or innovate around the incumbent's infrastructure.
- Cost advantage: If the original firm is genuinely most efficient, it stays in business by consistently offering the best combination of price and quality—otherwise, it risks losing its market share.

In short, under free entry and exit, a "natural monopoly" isn't inherently bad or exploitative; it's simply a scenario where a single firm's large-scale operations beat any competitor's smaller-scale costs. Government-imposed monopolies, on the other hand, lock out competition, leading to higher prices and poorer service. True market freedom ensures any so-called natural monopolist can't abuse consumers without inviting competitors—thus, it must keep prices and quality competitive on its own.

4 What causes Monopolies = State

4.1 Minimum Efficient Scale

A key factor that determines whether an industry ends up competitive or monopolistic is the minimum efficient scale (MES): the level of output at which average cost (AC) is minimized. If the MES is small compared to overall market demand, there is "room" for multiple firms to produce at (or near) their own efficient scale. In that situation, competitive conditions tend to prevail because each firm's share of the market is relatively modest, and no single producer can easily dominate the entire market.

However, if the MES is large relative to market demand, only one (or very few) firms can profitably operate at that efficient scale. This can lead to a monopoly or a highly concentrated market structure. The single firm (or small handful of firms) that captures the market can potentially restrict output and charge higher prices, resulting in deadweight loss and inefficiency.

4.2 Government's Role in Market Size

Even if a firm's technology implies a large MES, policy choices can expand or reduce the effective size of the market. For example, **open trade** allows domestic firms to sell to (and face competition from) global markets, reducing a single firm's ability to control prices. Conversely, **restrictive trade policies** shrink the domestic market, making it easier for a single producer to become dominant and behave like a monopolist. Other factors that create monopolies include:

- **Collusion / Cartels:** Multiple firms may collude to restrict output and raise prices, acting collectively like a monopolist. Antitrust laws generally prohibit this behavior.
- **First-Mover Advantages:** If one firm enters early and invests heavily in capacity ("tooling up"), it might deter new entrants by threatening price cuts, maintaining a dominant position.
- **Legal or Government Barriers:** Licenses, patents, or exclusive franchises can also create monopoly power if the government limits entry into an industry.

4.3 Libertarian Perspective

From a free-market viewpoint, the more open and contestable the market is (i.e., free trade, no legal barriers, no collusion), the less likely a single firm can maintain a harmful monopoly. If a firm truly has the lowest average cost due to economies of scale (a large MES) but no regulatory barriers protect it, then it must keep prices and quality competitive—otherwise potential rivals will enter or innovate around it. **Thus, policy choices (like protectionist measures) often matter more for entrenched monopoly power than purely technological factors. (= Trump)**

Most monopolies that are not "natural" (i.e., not driven purely by economies of scale and cost structure) tend to arise because the state imposes barriers to entry or grants exclusive privileges. These can include things like **licensing** requirements, patents or **intellectual property rules** that are overly broad, **legal franchise**

monopolies, or **trade restrictions** that shield an incumbent from foreign competition. In a fully open market—absent these artificial constraints—any firm charging monopoly prices would eventually face new rivals, unless its cost advantage was genuinely "natural." Hence, libertarians often argue that government action is the primary reason non-natural monopolies persist.

4.4 Example

Let's consider a numerical example showing how to:

- Compute the minimum efficient scale (MES) — the output level that minimizes average cost.
- **Compare MES to market demand** to judge whether the market can accommodate multiple efficient firms (suggesting competition) or only one (suggesting a natural tendency to monopoly).

Let's use the following setup. Suppose a firm's total cost function is

$$C(q) = 100 + 5q + q^2,$$

where:

- 100 is the fixed cost,
- $5q$ is a linear variable cost, and
- q^2 represents rising marginal costs as output grows.

The average cost (AC) is

$$AC(q) = \frac{C(q)}{q} = \frac{100 + 5q + q^2}{q} = \frac{100}{q} + 5 + q.$$

The MES is the output q that minimizes $AC(q)$. Differentiating $AC(q)$ w.r.t. q and setting to zero leads to:

$$\frac{d}{dq}\left[\frac{100}{q} + 5 + q\right] = -\frac{100}{q^2} + 1 = 0 \quad \longleftrightarrow \quad 1 = \frac{100}{q^2} \quad \longleftrightarrow \quad q^2 = 100 \quad \longleftrightarrow \quad q = 10.$$

Thus, the minimum efficient scale is $\boxed{q_{\text{MES}} = 10}$, and plugging $q = 10$ into $AC(q)$ gives the minimum average cost:

$$AC(10) = \frac{100}{10} + 5 + 10 = 10 + 5 + 10 = 25.$$

So the firm's average cost is lowest at \$25 per unit, when it produces 10 units. Let's consider 2 different cases to illustrate different market scenarios:

- **Case A:** Large Market Demand (e.g., $Q_D = 40$)
  - Total market demand is 40 units.
  - If one firm produces 10 units at the MES, the market still needs 30 more units to meet demand. In principle, up to four firms could each produce around 10 units at an efficient scale, or there might be even more firms producing slightly above/below 10 units.
  - Because several firms can operate near their MES, the industry can support multiple efficient producers, suggesting a more competitive market structure.
- **Case B:** Small Market Demand (e.g., $Q_D = 12$)
  - Total market demand is 12 units.
  - If one firm produces 10 units (its MES), that leaves only 2 units for any other potential entrant. Producing just 2 units is far below that entrant's own MES of 10, so its average cost would be much higher than 25, making it uncompetitive.

- In practice, only one firm can efficiently supply most or all of the 12-unit market. Other firms, if they enter at tiny output levels, face high average costs and would lose money. This can lead to a monopolistic or near-monopolistic market structure.

Clearly, there is a critical government influence on the market size. Even if technology dictates $q_{\text{MES}} = 10$, the effective size of the market depends on policy:

- Open Trade Policies: If the firm can sell to (and face competition from) foreign markets, total demand might be much larger than 40 or 12, reducing the chance of monopoly.
- Restrictive Trade Policies: If the firm can sell only domestically, total demand might be too small to support multiple efficient firms, effectively encouraging a monopolistic outcome.