lecture 2 – natural monopoly regulation
outline

- Natural monopoly
  - Definitions: economies of scale, economies of scope, subadditivity
  
- Regulation
  - Optimal solutions:
    - Linear and nonlinear pricing
    - Ramsey pricing
  
- Regulation in practice:
  - Rate of return regulation
  - Price caps
References

- Natural monopoly:
  - VHM, ch. 11
Natural monopoly
typical example

Let $C(q_i) = F + c q_i$. Then $AC_i = (F/q_i) + c$ is decreasing.
Natural monopoly (NM)
definition

- (cost-based or technology definition) An industry is a natural monopoly (NM) if the production of a particular good or service (or all combinations of outputs, in the multiple output case) by a single firm minimizes cost:
  - NM has been simply defined as existing when the AC curve is everywhere downward-sloping relative to market demand (economies of scale)
  - (Baumol et al., 1970) introduced formally the notion of subadditive costs; a NM occurs when the cost function is subadditive

- Tirole’s definition does not depend solely on costs: a NM arises when market equilibrium yields a single firm
Natural monopoly (NM) definition

- (cost-based or technology definition) An industry is a natural monopoly (NM) if the production of a particular good or service (or all combinations of outputs, in the multiple output case) by a single firm minimizes cost.
  - NM has been simply defined as existing when the AC curve is everywhere downward-sloping relative to market demand (*economies of scale*).
  - (Baumol et al., 1970) introduced formally the notion of subadditive costs; a NM occurs when the cost function is subadditive.

- Tirole’s definition does not depend solely on costs: a NM arises when market equilibrium yields a single firm.
Natural monopoly (NM) definition

- (cost-based or technology definition) An industry is a natural monopoly (NM) if the production of a particular good or service (or all combinations of outputs, in the multiple output case) by a single firm minimizes cost
  - NM has been simply defined as existing when the AC curve is everywhere downward-sloping relative to market demand (economies of scale)
  - (Baumol et al., 1970) introduced formally the notion of subadditive costs; a NM occurs when the cost function is subadditive

- Tirole’s definition does not depend solely on costs: a NM arises when market equilibrium yields a single firm
Economies of scale

- **Definition:** decreasing average long run cost as output increases

- **Why:**
  - Existence of substantial fixed costs
  - Opportunities for specialization in the deployment of resources
  - Strong market position in factor inputs
Economies of scale

![Graph showing Economies of Scale with LRAC and LRMC curves.](image)
Economies of scale with multiple outputs

Definitions (Baumol, Panzar, Willig):

1. **Decreasing AC along a ray:**
   \[ C(tQ) < tC(Q), \ t > 1 \]

2. **Decreasing average incremental cost:**
   \[ \frac{|C(q_1,q_2) - C(0,q_2)|}{q_1} \text{ decreasing with } q_1 \]

3. **Convex cost function along a transversal ray:**
   \[ C(tq_1,(1-t)q_2) < C(tq_1) + C((1-t)q_2) \]
   (similar to economies of scope – it’s cheaper to produce a convex combination of two goods in the same firm)
Subadditivity

definition

○ In a market with k firms, where firm i has a cost function $C(q_i)$ and total output is $Q$, firms’ cost functions are said to be subadditive at output level $Q$ when:

$$C(Q) < C(q_1) + C(q_2) + ... + C(q_k)$$

○ If this occurs for all values of $Q$, consistent with demand $Q = D(p)$, then the cost function is said to be globally subadditive
Subadditivity and economies of scale
single-product case

• In the single product case, economies of scale up to $q_i=Q$ is a sufficient but not a necessary condition for subadditivity over this range or, by the cost-based definition, for $NM^*$.

• In fact, it may still be less costly for output to be produced in a single firm rather than multiple firms even if output of a single firm has expanded beyond the point where there are economies of scale.
Subadditivity and economies of scale

One firm

Two firms

Fig. 11.4 VWH

Economies of scale
Subadditivity
Economies of scope

- Most NM (public utilities) produce more than one product and there is interdependence among outputs.

- Economies of scope exist when it is cheaper to produce two products together (joint production) than to produce them separately:

  \[ C(Q_1, Q_2) < C(Q_1, 0) + C(0, Q_2) \]

- Sources:
  - shared inputs
  - shared advertising creating a brand name
  - cost complementarities (producing one good reduces the cost of producing another)
Subadditivity and economies of scope

multiproduct case

- Economies of scope is a necessary but not sufficient condition for subadditivity

- In the multiproduct case, the existence of (product-specific) economies of scale in the production of any one product is neither necessary nor sufficient for subadditivity (because of economies of scope)

- Sufficient conditions for subadditivity:
  - economies of scope + declining average incremental cost for all products
  - Decreasing AC along a ray + convexity along a transversal ray
Natural monopoly
conflict: productive eff. vs. allocative eff.

- Is a NM productive-efficient?
  - Usually yes, but not always: Productive efficiency requires cost to be minimized

- Is a NM allocative-efficient?
  - No: A monopolist generates a deadweight loss by restricting output below the competitive level, since $P_M > MC$
Natural monopoly

efficiency

1. \((Q_e, P_e)\) first-best: \(P = MC\)
2. \((Q_0, P_0)\) second-best: \(P = AC\)
Natural monopoly

- Policy dilemma...

- Least-cost production requires a single-firm; but this leads to monopoly pricing – allocative inefficiency.

- Otherwise, competition results in productive inefficiency.
Natural monopoly

- Two-stage game
  - First stage: firms decide to enter (entry implies sunk cost of $k$)
  - Second stage: competition in prices
  - Unique pure strategy equilibrium: a single firm enters and sets $P = P_M$ (earning monopoly profit – $k$)
Natural monopoly solutions

- Doing nothing – why? Second-best obtained because of:
  - Contestable markets
Contestable markets

- Even if there are just a few firms in the market, there may be potential competition from firms who may enter the market.

- This may lead to the second best pricing solution!
Contestable markets

- Let there be N firms, of which m are producing

- The production vector is admissible iff there is market equilibrium and firms do not have losses

- The production vector is sustainable iff none of the N-m firms can enter the market with a lower price and have positive profit

- If a production vector is admissible + sustainable, then it’s contestable
Natural monopoly
solutions

- Doing nothing – why? Second-best obtained because of:
  - Contestable markets
  - Auction bidding
  - Close substitutes for the product

- Regulation – *ideal* pricing solutions
  - Linear pricing
    - Marginal cost pricing
    - Average cost pricing
  - Non linear pricing or multipart tariff
  - Ramsey pricing (multiproduct case)
Marginal cost pricing

Efficient MC price: \( P_0 = C' (Q(P_0)) \)

Advantage: allocative efficiency

Problems:
- information needed
- weak incentives to reduce costs
- NM is not able to break-even when economies of scale exist; use subsidy? This would imply raising funds (distortion) and the producer would know revenue gap would always be funded!
Moreover, we may have \( CS < TC \)
Average cost pricing

Efficient AC price: \( P_0 = \frac{C(Q(P_0))}{Q(P_0)} \)

Advantage: maximizes total welfare s.t. break-even constraint

Problems:
- information needed
- failure of allocative efficiency: less quantity and higher price than in MC pricing case (Deadweight loss)
- weak incentives to reduce costs
Nonlinear pricing

two-part tariffs

- Two-part tariffs include a fixed fee, regardless of consumption, plus a marginal cost price per unit
  \[ T(q) = A + Pq \]

- If \( P = c \), we may have efficient pricing and \( TR = TC \) for appropriate \( A \! \)

- Nonlinear pricing is more efficient than linear tariffs
  Often used in the utility industries (telecom., gas, water, electricity)
Nonlinear pricing

two-part tariffs

- If $C(q) = K + cq$ and consumers are homogeneous, then it would be optimal to set a two-part tariff with
  
  \[ A^* = \frac{K}{N} \text{ and } P^* = c \]

- But when consumers are heterogeneous, consumers with low willingness to pay drop out of the market if
  
  \[ K/N > CS(c) \]

- When consumers are heterogeneous, welfare maximizing nonlinear tariffs will most likely involve the firm offering consumers discriminatory two-part tariffs:
  
  - Quantity discounts
  - Multipart tariffs
  - Self-selecting tariffs
  
  (but discrimination may be forbidden....)
Nonlinear pricing
Increasing and declining block tariffs

Increasing block rate

Decreasing block rate

€

kWh

0.50
0.40
0.30
0.20

0.50
0.40
0.30
0.20

100 200 300

100 200 300
Nonlinear pricing
Multi-part tariff or self-selecting two-part tariffs

![Graph showing nonlinear pricing]

- **Total Expenditure (€)**
  - A: €5
  - B: €10
  - C: €20

- **Calls/month**
  - 100
  - 200

Legend:
- A: Base cost
- B: Additional cost for the second 100 calls
- C: Total cost for 200 calls
- D: No further increase in cost for additional calls
Nonlinear pricing
optimal two-part tariff

- **Trade-off:**
  - Efficiency losses because of exclusion of additional consumers when A raises
  - Consumption losses as P increases above marginal cost
  - (start with A=0 and P=c: the loss must be compensated by higher A or P or both; balance efficiency losses (consumer exclusion) with consumption losses (reduction quantity))

- **Optimal two-part tariffs generally involve a P that exceeds MC (no allocative efficiency) and a fixed fee that excludes some consumers from the market (failure of universal service)**
Multiproduct NM

- For multiproduct natural monopolist, MC pricing leads to negative profits.

- But if price for each product exceeds MC it can cover this shortfall,

- By how much?

  - In the context of a multiproduct monopolist, each product would have a linear price, and the set of prices would minimize deadweight social losses subject to the zero profit constraint.
The Ramsey rule

- The Ramsey rule or Ramsey-Boiteux pricing applies to multiproduct NM that would obtain losses with MC pricing.

- Ramsey found the result before (1927) in the context of the theory of taxation. The rule was later applied by M. Boiteux (1956) to NM.

- Ramsey prices are linear prices that satisfy zero profit and maximize social welfare.
The Ramsey rule

- Assumptions:
  - natural monopoly
  - independent demands (0 cross-price elasticities)
  - linear demands

- Ramsey-Boiteux pricing: the markup of each commodity is inversely proportional to the corresponding elasticity of demand (but it is smaller as the inverse elasticity of demand is multiplied by a constant lower than 1)

\[
\frac{P_i - MC_i}{P_i} = \frac{\lambda}{\varepsilon_i}
\]

- The rule implies that the relative change in quantity is the same for all goods
The Ramsey rule

example

- \( C(X,Y) = 1800 + 20X + 20Y \)

- Demands:
  - \( Q_x = 100 - P_x \)
  - \( Q_y = 120 - 2P_y \)

- MC pricing would imply \( P_x = P_y = 20 \); however, this implies losses

- One way is to increase the two prices proportionally until 36.1; this leads to DWL of 130 + 260 = 390

- An alternative is to raise the price of \( X \) (less elastic) more
The Ramsey rule

Proportionate price increase

Ramsey prices
Examples

- Rail rates for shipping sand, potatoes or oranges are lower than those for liquor, cigarettes,... because elasticities of demand of shipping products that have low values per pound are higher.

- But, before 1984, even though the elasticity of long-distance calls was higher than for short-distance calls (0.5-2.5 vs. 0.05-0.2), AT&T priced short-distance calls way below long-distance! Profits in long-distance were used to subsidize losses on local service.