

# 2

## Firm Heterogeneity and Competition

---

Pure economic profits in the macroeconomy can come from three main factors:

1. **Differentiation:** Product differentiation produces frictions in consumer substitution of goods. This is the most common feature in macroeconomic models used in policy by the BoE, NY Fed or ECB. (this chapter)
2. **Strategic interaction:** When markets are sufficiently small, entry barriers (such as fixed costs or regulation), or market size can allow firms to interact strategically to result in economic profit (next chapter).
3. **Technological advantages:** Innovation can give firms production advantages over competitors (last chapter).

### 2.1 Dixit and Stiglitz (1977) Monopolistic Competition

The model in Dixit and Stiglitz (1977) uses product differentiation and monopoly to create market power dynamics. As we will see, heterogeneity comes from differences in productivity (thus size) and activity.

#### 2.1.1 Consumption

We assume there are a continuum of goods indexed  $C_i$ . The household maximises consumption  $C$ , a CES aggregate of different goods  $C_i; i \in [0, 1]$  subject to their budget constraint,

$$\max C = \left( \int_0^1 C_i^{\frac{\varepsilon-1}{\varepsilon}} di \right)^{\frac{\varepsilon}{\varepsilon-1}} \quad (2.1)$$

$$\text{s.t. } W = \int_0^1 P_i C_i di \quad (2.2)$$

where  $W \in \mathbb{R}_+$  represents total expenditure, or wealth.

**Lemma 2.1.** *The households demand function is given by*

$$C_i = \left( \frac{P_i}{P} \right)^{-\varepsilon} C \quad (2.3)$$

where  $P$  is the aggregate price index that satisfies  $PC = W$ .

*Proof.* The Lagrangian function associated with the households maximisation problem is

$$\mathcal{L}(C_i, \lambda) = \left( \int_0^1 C_i^{\frac{\varepsilon-1}{\varepsilon}} di \right)^{\frac{\varepsilon}{\varepsilon-1}} + \lambda \left( W - \int_0^1 P_i C_i di \right) \quad (2.4)$$

The first order conditions imply

$$\frac{\partial \mathcal{L}}{\partial C_i} = 0 \implies \frac{\varepsilon}{\varepsilon - 1} \left( \frac{C_i}{C} \right)^{-\frac{1}{\varepsilon}} = \lambda P_i \quad (2.5)$$

$$\frac{\partial \mathcal{L}}{\partial \lambda} = 0 \implies W = \int_0^1 P_i C_i \, di \quad (2.6)$$

Since the first condition holds for any  $i, j \in [0, 1]$ , we have

$$C_i = C_j \left( \frac{P_i}{P_j} \right)^{-\varepsilon} \quad (2.7)$$

Using this expression,

- substituting into  $C$  in (2.1) yields

$$C = C_j P_j^\varepsilon \left( \int_0^1 P_i^{1-\varepsilon} \, di \right)^{\frac{\varepsilon}{\varepsilon-1}} \quad (2.8)$$

- substituting into  $W$  in (2.2) yields

$$W = C_j P_j^\varepsilon \int_0^1 P_i^{1-\varepsilon} \, di \quad (2.9)$$

To proceed we introduce the price index through the following claim.

**Claim.** The price index,  $P$  defined as  $PC = W$ , can be shown to equal

$$P = \left( \int_0^1 P_i^{1-\varepsilon} \, di \right)^{\frac{1}{1-\varepsilon}} \quad (2.10)$$

*Proof of Claim.* Divide (2.2) by (2.1) to yield  $P$ . □

Using this result, we can rewrite (2.8) as

$$C_{jt} = \left( \frac{P_j}{P} \right)^{-\varepsilon} C \quad (2.11)$$

Since  $j \in [0, 1]$  was arbitrary, it holds for all goods. □

## 2.1.2 Production

Firms are monopolies in their respective sectors  $i \in [0, 1]$ . They produce with the linear production function

$$Y_i = A_i L_i \quad (2.12)$$

where  $A_i$  represents an exogenous productivity variable. Firms face a cost function  $\Psi : \mathbb{R} \rightarrow \mathbb{R}$  given by

$$\Psi(L_i) = w_i L_i \quad (2.13)$$

**Lemma 2.2.** The firms marginal cost,  $\psi_i := \frac{\partial}{\partial Y_i} \Psi$  is given by

$$\psi_i = \frac{w_i}{A_i} \quad (2.14)$$

*Proof.* By the definition of marginal cost and the chain rule,

$$\frac{\partial \Psi}{\partial Y_i} = \frac{\partial \Psi}{\partial L_i} \frac{\partial L_i}{\partial Y_i} \quad (2.15)$$

$$= \frac{\partial \Psi}{\partial L_i} \left( \frac{\partial Y_i}{\partial L_i} \right)^{-1} \quad (2.16)$$

The results follows clearly from the last line. □

### 2.1.3 Equilibrium pricing and markups

**Definition 2.3** (Markup). A markup,  $\mu$  is the ratio between the price sold for a good, and the marginal cost  $\psi_i$  to produce that good. The markup satisfies

$$P_i = \mu_i \psi_i \quad (2.17)$$

**Definition 2.4** (Competitive Equilibrium). In *equilibrium*, markets clear for all firms and as a whole, i.e. for all  $i \in [0, 1]$ ,  $C_i = Y_i$  and  $C = Y$ . In particular, a *competitive equilibrium* is when households and firms optimise their consumption and production decisions.

If firms maximise profit, the model implies an endogenous markup as characterised in the following theorem.

**Theorem 2.5.** *Optimal price setting by firms in the Dixit and Stiglitz model implies endogenous markups are given by*

$$\mu_i = \frac{\varepsilon}{\varepsilon - 1} \quad (2.18)$$

Note that markups are homogeneous, i.e. independent of  $i$ .

*Proof.* Profit is revenue minus cost; firm maximises profit subject to the demand function in (2.3) and the production function in (2.12), by choosing their price:

$$\max_{P_i} \Pi_i = P_i Y_i - \Psi(L_i) \quad (2.19)$$

$$\text{s.t. } Y_i = \left(\frac{P_i}{P}\right)^{-\varepsilon} Y, \quad L_i = \frac{Y_i}{A_i} \quad (2.20)$$

Substituting the constraints into the objective function yields an unconstrained problem; taking first order conditions of this yields

$$\frac{\partial \Pi}{\partial P_i} = 0 \implies (1 - \varepsilon) Y_i = \frac{\partial \Psi(L_i)}{\partial P_i} \quad (2.21)$$

By the chain rule, the right hand side can be broken into

$$\frac{\partial \Psi}{\partial Y_i} \frac{\partial Y_i}{\partial P_i} = -\varepsilon \psi P_i^{-1} Y_i \quad (2.22)$$

Combining these results and rearranging yields the result.  $\square$

**Corollary 2.6.** *Under perfect competition, there is little product diversification, i.e.  $\varepsilon \rightarrow \infty$ , and so there are no markups, i.e.  $\mu = 1$ .*

*Proof.* From Theorem 2.5, we take the limit of  $\mu_i$  as  $\varepsilon \rightarrow \infty$ :

$$\begin{aligned} \lim_{\varepsilon \rightarrow \infty} \mu_i &= \lim_{\varepsilon \rightarrow \infty} \frac{\varepsilon}{\varepsilon - 1} \\ &= \lim_{\varepsilon \rightarrow \infty} \left(1 + \frac{1}{\varepsilon - 1}\right) \\ &= 1 \end{aligned}$$

$\square$

### 2.1.4 Concentration and markups

High productivity means a firm has low marginal costs. Using the optimal markup by firms, we can derive a relationship between markups and concentration.

**Theorem 2.7.** *Firm concentration is decreasing in markups.*

*Proof.* Consider two firms  $i$  and  $j$ . Their relative demand as given in (2.7) is

$$\frac{C_i}{C_j} = \left( \frac{P_i}{P_j} \right)^{-\varepsilon} \quad (2.23)$$

As markups are homogeneous and equal to  $\frac{\varepsilon}{\varepsilon-1}$ , we have

$$\frac{C_i}{C_j} = \left( \frac{\psi_i}{\psi_j} \right)^{\frac{\mu}{1-\mu}} \quad (2.24)$$

As  $\mu \rightarrow \infty$ , the exponent  $\frac{\mu}{1-\mu} \uparrow -1$ , since we know  $\mu \geq 1$ . When the exponent is  $-1$ , the ratio of consumption is minimised.  $\square$

### 2.1.5 Labour markets and markups

Using the definition of markups, we can rewrite optimal consumption for goods by firm  $i$  as

$$C_i = \left( \frac{\mu_i \psi_i}{P} \right)^{-\varepsilon} C \quad (2.25)$$

This implies

- higher productivity  $A_i$ , increases consumption,
- higher markups  $\mu_i$ , reduce consumption
- the elasticity of substitution  $\varepsilon$ , determines the sensitivity of demand to prices.

In particular, using the expression for marginal costs, we can relate wages to markups.

**Theorem 2.8.** *If nominal wages and markups are homogeneous, i.e. for all  $i \in [0, 1]$ ,  $\mu_i = \mu$  and  $w_i = w$ , then real wages in a competitive equilibrium are given by*

$$\frac{w}{P} = \frac{1}{\mu} \left( \int_0^1 A_i^{\varepsilon-1} di \right)^{\frac{1}{\varepsilon-1}} \quad (2.26)$$

*Proof.* Substitute optimal consumption (2.3) for  $C_i$  into the definition of aggregate consumption  $C$  in (2.1) to yield

$$C = \left( \int_0^1 \left[ \left( \frac{\mu \psi_i}{P} \right)^{-\varepsilon} C \right]^{\frac{\varepsilon-1}{\varepsilon}} di \right)^{\frac{\varepsilon}{\varepsilon-1}} \quad (2.27)$$

$$= C \left( \int_0^1 \left( \frac{\mu w}{A_i P} \right)^{1-\varepsilon} di \right)^{\frac{\varepsilon}{\varepsilon-1}} \quad (2.28)$$

As nominal wages and markups are assumed constant, we can extract them from inside the integral. Rearranging for  $w/P$  yields the result.  $\square$

Although Theorem 2.8 imposes homogeneity of markups, as we showed in Theorem 2.5, markups in the Dixit and Stiglitz model are in fact constant. Therefore the only additional assumption imposed is on the homogeneity of wages. These assumptions also make further predictions about labour supply and its share of income in the Dixit and Stiglitz (1977) economy. To see this we first define *labour supply*.

**Definition 2.9** (Labour supply). *Labour supply*,  $L$  is the sum of all intermediate labour demanded by firms,

$$L := \int_0^1 L_i \, di \quad (2.29)$$

**Lemma 2.10.** *In a competitive equilibrium with homogenous wages and markups, labor supply is related to aggregate consumption by*

$$L = \left( \int_0^1 A_i^{\varepsilon-1} \, di \right)^{\frac{1}{1-\varepsilon}} Y \quad (2.30)$$

*Proof.* In equilibrium,  $C_i = Y_i = A_i L_i$ . Therefore labour supply can be written as

$$L = \int_0^1 \frac{Y_i}{A_i} \, di \quad (2.31)$$

$$= \int_0^1 A_i^{\varepsilon-1} \left( \frac{\mu w}{P} \right)^{-\varepsilon} Y \, di \quad (2.32)$$

where the second line comes from substituting (2.25) in for  $C_i = Y_i$ . Since real wages are given by ( ), substituting in and rearranging yields the result.  $\square$

**Theorem 2.11.** *In a competitive equilibrium, the share of income given to labour relative to total wealth is given by*

$$\frac{wL}{PY} = \frac{1}{\mu} \quad (2.33)$$

*Thus labour income is decreasing in markups,  $\mu$ .*

*Proof.* From (2.26) and (2.30) we have

$$wL = \overbrace{\frac{P}{\mu} \left( \int_0^1 A_i^{\varepsilon-1} \, di \right)^{\frac{1}{\varepsilon-1}}}^w \overbrace{\left( \int_0^1 A_i^{\varepsilon-1} \, di \right)^{\frac{1}{1-\varepsilon}} Y}^L \quad (2.34)$$

$$= \frac{PY}{\mu} \quad (2.35)$$

Dividing by  $PY$  leaves the result.  $\square$

How should we interpret this result? Higher markups cause a smaller fraction of income to go to factors (labour). The mechanism as we looked at above is through reducing labour demand, and so lowering wage for labour. We say that markups cause a *wedge* between wages and the marginal product of labour. This result matches the empirical observations from Chapter 1; the stylised facts presented show a continuous decrease in the US labour share of income (see ??; Van Reenen 2018), whilst markups have steadily risen as a whole over the past few decades (see ??; De Loecker et al. 2020).

## 2.2 Measuring Markups

TLDR: measuring markups is difficult. Markups are the ratio of prices and marginal costs, which are both not included in accounting data. Therefore we need to come up with a suitable estimator to obtain firm level markups from accounting data.

### 2.2.1 Estimating the production function

Consider the Cobb-Douglas production function

$$Y_{it} = A_{it} K_{it}^{\alpha} L_{it}^{\beta} M_{it}^{\gamma} \quad (2.36)$$

and its log transform

$$y_{it} = a_{it} + \alpha k_{it} + \beta l_{it} + \gamma m_{it} \quad (2.37)$$

If we assume log productivity follows an AR(1) process,

$$a_{it} = \rho a_{i,t-1} + \varepsilon_{it} \quad (2.38)$$

**Identification Assumption.** Firms do not know  $\varepsilon_{it}$  at  $t - 1$ . Therefore  $k_{i,t-1}$ ,  $l_{i,t-1}$  and  $m_{i,t-1}$  uncorrelated with  $\varepsilon_{it}$ . Using this assumption, Akerberg et al. (2015) implement the process

1. guess values for  $\alpha, \beta, \gamma$ ,
2. compute the implied productivity time series with the guesses through

$$a_{it} = y_{it} - \alpha k_{it} + \beta l_{it} + \gamma m_{it} \quad (2.39)$$

3. estimate the auto-regression on  $a_{it}$ ,
4. update the guesses for  $\alpha, \beta, \gamma$ , until  $\varepsilon_{it}$  is not correlated with the factors.

### 2.2.2 The Hall (1988) markup estimator

Hall (1988) proposes the following markup estimator.

**Definition 2.12.** Hall (1988) estimates markups using

$$\mu_{it} = \gamma \frac{P_{it} Y_{it}}{p_t^{(m)} M_{it}} \quad (2.40)$$

where the right hand side is the ratio of sales to intermediary input prices multiplied by the income share.

*Derivation.* Consider a firm with the standard Cobb-Douglas production function:

$$Y_{it} = A_{it} K_{it}^{\alpha} L_{it}^{\beta} M_{it}^{\gamma} \quad (2.41)$$

where  $K$  is capital,  $L$  is labour and  $M$  is intermediary inputs. The firms objective is to minimise costs for a given output level, i.e.

$$\min \Psi_{it} = r_t K_{it} + w_t L_{it} + p_t^{(m)} M_{it} \quad (2.42)$$

$$Y = A_{it} K_{it}^{\alpha} L_{it}^{\beta} M_{it}^{\gamma}; \quad Y \in \mathbb{R}_+ \text{ given} \quad (2.43)$$

with aggregate factor prices  $r_t, w_t$  and  $p_t^{(m)}$  corresponding to capital, labour and intermediates respectively. The Lagrangian associated with the firm's problem is

$$\mathcal{L} = r_t K_{it} + w_t L_{it} + p_t^{(m)} M_{it} + \lambda (Y - A_{it} K_{it}^\alpha L_{it}^\beta M_{it}^\gamma) \quad (2.44)$$

which has the first order condition with respect to  $M_{it}$ ,

$$\frac{\partial \mathcal{L}}{\partial M_{it}} = 0 \implies p_t^{(m)} = \lambda \gamma Y_{it} M_{it}^{-1} \quad (2.45)$$

Note the definition of the Lagrange multiplier here; it represents the change in the objective function, for a unit increase in the constraint, i.e.

$$\lambda = \frac{\partial \Psi_{it}}{\partial Y_{it}} = \psi_{it} \quad (2.46)$$

which in this context is just marginal cost,  $\psi_{it}$ . By the definition of a markup in (2.17), they are related to marginal costs through  $\psi_{it} = P_{it}/\mu_{it}$ . Substituting this into (2.45) and rearranging yields the desired expression.  $\square$