Chapter 21
A Macroeconomic Model of Monopolistic Competition:
The Dixit-Stiglitz Framework

The RBC view of the macroeconomy is premised on perfect competition in all three macro markets (goods markets, labor markets, and financial markets). For the seminal issue of the degree of (goods) price stickiness, it is goods markets on which we need to focus, so we limit our attention to goods markets from here on.

In perfect competition, there is a sense in which no supplier makes any purposeful, meaningful decision regarding the price that *it* sets. Rather, because of perfect substitutability between all products (recall the assumption of *homogenous goods* in a perfectly-competitive market), firms are all *price-takers*. A view of firms as price-takers is incompatible with the notion that we would now like to entertain, that of *firms only infrequently setting their prices*. Thus, the most basic step we must take in order to even begin to conceptually understand the idea of (possibly sticky) price-setting is to assert that firms are indeed price-*setters*, rather than pure price-*takers.*

As you should recall from basic microeconomics, the market structure of *monopoly* offers a relatively easy analytical framework in which firms are indeed price-*setters*. However, from the point of view of macroeconomics, pure monopoly seems an untenable view to adopt. After all, it is implausible, at the aggregate level, to asset that there is *one* producer of *all* of the goods that are produced and sold in the economy. A more realistic view should admit the simple fact that there are many producers of goods as well as the fact that these goods are not all identical to each other. That is, there is some *imperfect substitutability* between the many goods an economy produces.

The concept of **monopolistic competition** offers an intermediate theoretical ground between pure monopoly and perfect competition. Indeed, the terminology itself suggests that the concept is an intermediate one between pure monopoly and perfect competition. Modern New Keynesian models are based on a monopolistically-competitive view of goods markets, in contrast to the RBC framework's perfectly-competitive view. The basic economic idea underlying a monopolistically-competitive view of goods markets is that there are many goods that consumers purchase and that they all are, to some degree, imperfect substitutes for each other.

In what follows, we will lay out the basic theoretical structure of macroeconomic models based on monopolistic competition. Before beginning, though, we define an important concept for the analysis of models employing or based on monopolistic competition.

Markup

We will often want to speak of by how much a firm's (presumably, optimally-chosen) chosen price, on a per-unit basis, exceeds the cost of production of a given unit of the good. As you should recall from basic microeconomics, a firm's cost of producing a given (i.e., the marginal) unit of output is measured by its **marginal cost.**

A firm's **gross markup** is defined as the (per-unit) price it charges divided by its marginal cost. Denoting by p the unit price chosen by a firm, by mc the firm's marginal cost of production, and by μ, we thus have that

$$\mu = \frac{p}{mc}.$$

Recall from basic microeconomics that *in a perfectly-competitive market, market forces dictate that* $p = mc$. Thus, we have that $\mu = 1$ in a perfectly-competitive market.[175] The interpretation of this is that a firm operating under the conditions of perfect competition has no scope whatsoever to earn a (marginal) profit on the goods it sells. Again recalling results and ideas from basic microeconomics, zero marginal profits is consistent with the idea that in perfect competition, firms earn zero (economic, as distinct from accounting) total profits.

As we will see below, a firm operating in a monopolistically-competitive market will earn positive (marginal) profits, and thus will be able to achieve a gross markup of $\mu > 1$.

Retail Firms

From an aggregate perspective, monopolistic competition forces us, among other things, to confront the fact consumers purchase a wide variety of goods. For theoretical modeling purposes, however, it turns out to be convenient to assume a structure in which consumers purchase just *one* (type of) good, just as in the RBC view we have adopted thus far. Thus, we will continue using the concept of the "consumption basket"

[175] We can also define the concept of a firm's **net markup,** which is the percentage by which price exceeds marginal cost. In the case of perfect competition, clearly the net markup is zero percent. For many applications, gross markup is an easier concept with which to work, so we will almost solely rely on it rather than net markup.

purchased by the representative consumer (i.e., we will still be able to speak of "all stuff" consumption). However, we will slightly relabel some of our concepts.

We will call the (homogenous) good (the consumption basket) that consumers purchase **retail goods.** Retail goods are assumed to be sold by **retail-goods producing firms** in a perfectly competitive market. That is, we will assume that a given retail firm is completely identical in every respect, including in what good it sells, to every other retail firm. The implication of this is that we can suppose that there is a representative retail firm.

Denote by y_t the quantity of retail good that the representative retail firm sells, and by P_t the nominal price of a unit of retail good. Because we are assuming that retailers sell their output in a perfectly competitive goods market, there thus far is nothing different, apart from some relabeling of concepts, from the RBC-style view we have adopted up until now.

Here is where we layer in monopolistic competition. In order to produce the retail good, a retailer must purchase a great many **wholesale goods.** That is, the inputs into the "production process" of a retail firm are themselves goods.[176] As a heuristic, think of a large department store that purchases items (clothes, furniture, electronics, jewelry, etc.) from a great many manufacturers and puts them "on display" in its retail outlets. In this example, the "wholesale goods" would be the great many clothes, electronics, etc. that the retailer purchases, and the "retail good" is the "basket of goods" that the store offers to its customers.

How many is a "great many" wholesale goods? Casual introspection about the world suggests *a lot* of goods and services comprise the aggregate "consumption basket." While consumers do not face literally an *infinite* number of possible goods they can purchase, clearly the number is somewhat beyond our comprehension, especially when one takes into account the fact that there various sizes, colors, styles, etc. for many seemingly identical goods. For this reason and because it is convenient mathematically, we will assert that "many" means "infinite." Specifically, we will assume that there is a *continuum* of wholesale goods, and each good is indexed on the unit interval [0,1]. Thus, note that we will work with a *continuous* number of wholesale goods, rather than with a *discrete* number of goods.[177]

To be a bit more concrete, suppose that every point on the unit interval [0,1] represents a particular wholesale good. Each of these goods is imperceptible – infinitesimally small – when compared to the entire spectrum of goods available, which seems like a plausible representation of the reality described above. We will assume that each good that lies on

[176] For simplicity, we will abstract from other types of inputs (such as capital and labor) that retailers might require. That is, we are assuming that it is *only* wholesale goods that are required for the production of retail goods.

[177] Because applying the tools of calculus typically requires continuous, as opposed to discrete, objects.

the unit interval is produced by a unique **wholesale goods producer** and is *imperfectly substitutable* with any other of these goods. Thus, these goods that lie on the unit interval – these wholesale goods – are differentiated products, which, as we stated above, allows us to admit the possibility of some monopoly power. We will describe wholesale goods producers in the next section.

First, though, we must describe the "production technology" and profit maximization problem that retail goods firms solve. In very general terms, we can describe the activities in which a retail goods firm engages as the following: it must purchase (via markets) each of the wholesale goods, apply some "packaging" or "transformation" technology to them (i.e., provide "retail services" that allow consumers to purchase the final "consumption basket"), and then sell the resulting retail good.

Since the incorporation of the idea of monopolistic competition into mainstream macroeconomics in the 1980's and 1990's, the most commonly-employed functional specification for the "packaging technology" of retail firms is the **Dixit-Stiglitz aggregator,**

$$y_t = \left[\int_0^1 y_{it}^{1/\varepsilon} di \right]^{\varepsilon}.$$

In this expression, y_t is the output, in period t, of the retailers, and y_{it}, for $i \in [0,1]$ (note well the notation here), is wholesale good i, of which, recall, there is an infinite number.[178] The parameter ε measures the curvature of this aggregation (aka packaging, aka transformation) technology. Basic monopoly theory requires that $\varepsilon > 1$. In the limit, as $\varepsilon \to 1$, obviously we would have $y_t = \int_0^1 y_{it} di$. With $\varepsilon = 1$, the resulting *linear* aggregation technology implies that each of the wholesale goods are *perfect* substitutes for each other, which undermines our whole analytical objective.

In the context of our theoretical model, allowing for curvature (i.e., $\varepsilon > 1$) in the aggregation technology is the basis for the existence of monopolistic competition. What curvature achieves for us is that retail firms *must* purchase some of every type of wholesale good. To continue the department store example from above, this means that a retailer wants to purchase some TV's, some shirts, some pants, some watches, some men's shoes, and so on – it wants to have some of every type of product on hand for the customers that it sells to. As will become clear below when we study wholesale goods firms, the parameter ε will also denotes the gross markup that they (the wholesale goods firms) charge.

[178] See Dixit, Avinash K. and Joseph E. Stiglitz. 1977. "Monopolistic Competition and Optimum Product Diversity." *American Economic Review,* Vol. 67, p. 297-308.

Denote by P_t the nominal price of the retail good (i.e., the per-unit price of the retail good) and by P_{it} the nominal price of wholesale good i, $i \in [0,1]$. The price of any wholesale good is taken as given by the retail firm – thus, we assume that there are no "negotiations" between retail firms and wholesale firms.[179] The profit function of the representative retailer is thus

$$P_t y_t - \int_0^1 P_{it} y_{it} di ,$$

which is simply its total revenue net of its total costs (recall the assumption we made above that purchases of wholesale goods are the only cost item for a retailer). Inserting the aggregator technology from above, we can re-express the profit function as

$$P_t \left[\int_0^1 y_{it}^{1/\varepsilon} di \right]^{\varepsilon} - \int_0^1 P_{it} y_{it} di .$$

As we just stated, the retail firm takes as given the price P_{it} of any given wholesale good i. Because we have assumed that retail goods are sold to consumers in perfectly-competitive product markets, it also takes as given the price P_t of the (retail) goods that it sells. Hence, the only object(s) of choice in the above profit function are the individual y_{it}'s, for each $i \in [0,1]$. That is, given the input and output prices it faces, the retail firm makes an optimal choice with respect to each wholesale good in order to maximize its profits, which are given by the previous expression.

Let's focus on good j within the unit interval [0,1].[180] Taking the first-order condition of the profit function (with the Dixit-Stiglitz aggregator substituted in, as in the second expression of profits presented above) with respect to y_{jt}, we have

$$\varepsilon P_t \left[\int_0^1 y_{it}^{1/\varepsilon} di \right]^{\varepsilon - 1} \frac{1}{\varepsilon} y_{jt}^{\frac{1}{\varepsilon} - 1} - P_{jt} = 0 .$$

Note carefully how the first term of this first-order condition arises – it arises by use of the chain rule of calculus, realizing that differentiation can be performed underneath an integration, and being careful about the distinction between product j and the arbitrary index of integration i.[181] We can simplify this expression to a very useful and

[179] A more nuanced view of the world probably would want to admit that, because they both often can be large players, retailers and wholesalers do "negotiate" with each and, thus, neither necessarily needs to be thought of as a price-taker. This is a topic for a more advanced course in macroeconomic theory.

[180] Keep straight the use of the indexes i and j. In the integrals we have so far written down, i is a dummy index of integration – we know this from the fact that di appears in each integral we have written down. Thus, i is simply keeping track of goods as we "loop through" the integral; i in these integrals is not referring to any particular good within [0,1].

[181] A technical point you may recall from basic calculus is that integration and differentiation are both linear operators. Linear operators are commutative; hence their order of operations can be interchanged

interpretable form. In the several steps that follow, what we will do is rearrange this expression into an expression that is easily understandable as the **demand function for wholesale good *j***.

First, cancel the ε terms. Next, we will dramatically simplify the term in square brackets. To do this, note that, by appropriately manipulating exponents, the Dixit-Stiglitz aggregator can be re-written as

$$y_t^{1/\varepsilon} = \int_0^1 y_{it}^{1/\varepsilon} di .$$

Then, raising both sides of this expression to the power ε-1, we have

$$y_t^{\frac{\varepsilon-1}{\varepsilon}} = \left[\int_0^1 y_{it}^{1/\varepsilon} di\right]^{\varepsilon-1} .$$

The right-hand-side is exactly the term we wanted to eliminate from the retail firm's first-order condition. Now making this substitution, the retail firm's first-order condition can be expressed as

$$P_t y_t^{\frac{\varepsilon-1}{\varepsilon}} y_{jt}^{\frac{1}{\varepsilon}-1} = P_{jt} .$$

We want to now isolate the y_{jt} term. Combining exponents and rearranging, we have

$$y_{jt}^{\frac{1-\varepsilon}{\varepsilon}} = \left(\frac{P_{jt}}{P_t}\right) y_t^{\frac{1-\varepsilon}{\varepsilon}} .$$

Finally, raising both sides to the power $\varepsilon/(1-\varepsilon)$, we have

$$y_{jt} = \left(\frac{P_{jt}}{P_t}\right)^{\frac{\varepsilon}{1-\varepsilon}} y_t .$$

This expression is **the demand function for wholesale good *j*.** Note that it indeed has the basic properties required of any demand function: it articulates an inverse relationship between the price P_{jt} of wholesale good j and the demand for it, y_{jt} (holding everything else, namely P_t and y_t, constant).[182] This demand function is an important

freely. This property is what allows us to differentiate in a very straightforward way with respect to y_{jt} inside the integral.

[182] Recall from above that we must have, consistent with monopoly theory, $\varepsilon > 1$. This means that the exponent to which the term (P_{jt}/P_t) is raised is a negative number.

building block for our description of a wholesale goods firm's profit-maximization problem, which we describe next.

The demand function we have derived is for good j. Clearly, we would obtain an identical-looking demand function for any other wholesale good, say good k. That is, we could have started this entire analysis by taking the retail firm's first-order condition with respect to wholesale good k; apart from replacing j by k everywhere in our analysis, nothing would be substantively different.

Wholesale Firms

Now we turn to a description of the activities of wholesale firms. As described above, wholesale firms (of which, recall again, there is a continuum) sell their differentiated output to retailers in a perfectly-competitive market. Due to the differentiated nature of wholesale products, wholesale firms have some market power and thus are explicitly price-*setters*.

We make two additional auxiliary assumptions regarding wholesale firms (relaxing these would not substantively change the conclusions of our analysis; the expense of doing so is some more cumbersome mathematics). First, suppose that there are no fixed costs of production. As you should recall from basic microeconomics, this means that the average variable cost of production is equal to the average total cost of production. Second, suppose that the per-unit production cost of each unit of wholesale output is identical **regardless of the scale of production.** In the language of basic microeconomics, this means we are assuming that wholesale firms have production technologies that exhibit **constant returns to scale,** which has the consequence that the wholesale firm's *marginal cost of production is invariant to the quantity that it chooses to produce.*

Coupled together, these two assumptions lead to the mathematically convenient consequence that the marginal cost function coincides with the average total cost function. In turn, this means that total costs of production can be expressed simply as the quantity produced times the marginal cost of production.[183]

Let's continue to focus on the particular wholesale good j. Given our assumptions, the profit function of wholesale firm j in period t can be expressed as

$$P_{jt}y_{jt} - P_t mc_{jt} y_{jt},$$

[183] There is a lot of basic microeconomics underlying these results and conclusions. It's probably worthwhile to convince yourself or review that all this is correct from the point of view of first principles of microeconomics.

We have expressed profits in nominal terms. The first term $P_{jt}y_{jt}$, which is wholesaler *j*'s total revenues, is clearly in nominal terms. In the cost term, $mc_{jt}y_{jt}$ denotes the **real** total cost to wholesaler *j* of producing y_{jt} units of output. Thus, the way we will denote things is that *mc* measures the **real** (not the nominal) marginal cost of production.[184] To then turn this into a nominal object, we multiply by P_t, which is the economy-wide nominal price level, which, in our environment, is simply the price of the "bundled" retail good. Note carefully that we are multiplying by P_t, not by P_{jt}, to convert into nominal units here.

A monopolist (in our environment, monopolistic competitor *j*) **takes as given the demand function it faces** when making its profit-maximizing choices. This is where the demand function for wholesale good *j* that we derived above comes into use. Substituting in the demand function for good *j* (alternatively, we could express it as a constraint on the optimization problem and introduce a Lagrange multiplier), we can express wholesale firm *j*'s profit function as

$$P_{jt}\left(\frac{P_{jt}}{P_t}\right)^{\frac{\varepsilon}{1-\varepsilon}} y_t - P_t mc_t \left(\frac{P_{jt}}{P_t}\right)^{\frac{\varepsilon}{1-\varepsilon}} y_t .$$

This term looks quite messy at first glance, but the path of the rest of our analysis is now clear: the only object in this expression over which wholesaler *j* has any control is the price P_{jt} it charges (it does, after all, have monopoly power). Thus, the next step is to compute the first-order condition of profits with respect to P_{jt} and then analyze the resulting optimal price.

To make our algebra a bit more transparent, we can combine the relevant P_{jt} terms in the profit expression. We can combine the above to

$$P_{jt}^{\frac{1}{1-\varepsilon}} P_t^{\frac{\varepsilon}{\varepsilon-1}} y_t - P_{jt}^{\frac{\varepsilon}{1-\varepsilon}} P_t^{\frac{2\varepsilon-1}{\varepsilon-1}} mc_t y_t .$$

The first-order condition of this expression (which is still simply wholesaler *j*'s profits in period *t*) with respect to P_{jt} is

$$\frac{1}{1-\varepsilon} P_{jt}^{\frac{\varepsilon}{1-\varepsilon}} P_t^{\frac{\varepsilon}{\varepsilon-1}} y_t - \frac{\varepsilon}{1-\varepsilon} P_{jt}^{\frac{2\varepsilon-1}{1-\varepsilon}} P_t^{\frac{2\varepsilon-1}{\varepsilon-1}} mc_t y_t = 0 .$$

[184] We have also allowed for the wholesaler-specific index *j* in the specification of marginal cost, though we will not actually make use of this. That is, we will only consider cases in which *mc* is identical between any two wholesalers *j* and *k*, $j \neq k$. This is tantamount to assuming that not only does each wholesaler use a constant-returns-to-scale production technology, but that each wholesaler uses *the same* constant-returns-to-scale production technology. Thus, from here on, we drop the firm-specific index on marginal costs.

We can now obviously cancel the y_t terms as well as the 1/(1-ε) terms; doing so leaves us with the slightly easier expression

$$P_{jt}^{\frac{\varepsilon}{1-\varepsilon}} P_t^{\frac{\varepsilon}{\varepsilon-1}} - \varepsilon P_{jt}^{\frac{2\varepsilon-1}{1-\varepsilon}} P_t^{\frac{2\varepsilon-1}{\varepsilon-1}} mc_t = 0 .$$

Let's continue to compact this expression. Multiply the entire expression by $P_{jt}^{\frac{-\varepsilon}{1-\varepsilon}}$, which leaves us with

$$P_t^{\frac{\varepsilon}{\varepsilon-1}} - \varepsilon P_{jt}^{-1} P_t^{\frac{2\varepsilon-1}{\varepsilon-1}} mc_t = 0 .$$

Next, multiply this expression by $P_t^{\frac{-\varepsilon}{\varepsilon-1}}$, which leaves us with

$$1 - \varepsilon P_{jt}^{-1} P_t mc_t = 0 .$$

Finally, solving for the profit-maximizing price of wholesale good *j*,

$$P_{jt} = \varepsilon P_t mc_t .$$

If we define wholesale firm *j*'s **relative price** as $p_{jt} = P_{jt}/P_t$ (note the distinction between lowercase and uppercase notation!), which is the **real** (in units of retail goods – as opposed to in units of currency) price charged by wholesaler *j*, we can instead express the profit-maximizing price as

$$p_{jt} = \varepsilon mc_t .$$

Regardless of which way we prefer to view things (the profit-maximizing price chosen by wholesaler *j* in nominal terms or in real terms), a very important result emerges here: **the profit-maximizing choice is a simple markup over marginal cost. Moreover, the markup is time-invariant: regardless of any shocks hitting the economy, in *every* period *t*, the wholesale firm sets its price as a constant markup over marginal cost.** The markup is given by ε (which we noted above it would), which controls the curvature of our Dixit-Stiglitz aggregator. Recall that in a monopolistic (or monopolistically-competitive) environment, it must be the case that $\varepsilon > 1$. This means that wholesale firms here are earning positive marginal profits (and, indeed, due to our assumption of zero fixed costs, positive total profits, as well). Given our precise definition of gross markup, it is clear that the (optimal) markup here turns out to be

$$\mu = \frac{p_{jt}}{mc_t} = \varepsilon$$

in every period.

Discussion

As with any theoretical structure (whether in economics or any field), the pure Dixit-Stiglitz-based view, which implies (among other things) that firms never alter their markups, taken too literally is an untenable view of the world. Lots of empirical evidence suggests that firms *do* change their markups, sometimes in very specific, predictable ways (i.e., firms holding seasonal "sales" can be interpreted in terms of a strategic change in the markup that it changes). Even at the aggregate level, evidence suggests that markups fluctuate at business cycle frequencies.[185] A great deal of research attempts to uncover why markups fluctuate over the course of the business cycle, but there really is no compelling explanation (which is another way of saying that there are a great many possible explanations, but none of them so far has stood out as obviously *the* main reason).

Nevertheless, being quite tractable, the Dixit-Stiglitz structure has become ubiquitous as a foundation of modern New Keynesian models. All of our analysis thus far has presumed completely flexible prices – that is, we have nowhere asserted that price adjustment entails any "menu costs" or are "sticky" in any way. And yet, as we mentioned at the outset, a theory that asserts that firms only infrequently (re-)set their prices or incur costs of setting prices requires, as a prerequisite, adopting a view in which firms are price-*setters* in the first place. The Dixit-Stiglitz structure at least makes progress on this front. Next, we turn to the most basic sticky-price New Keynesian model based on the Dixit-Stiglitz structure.

[185] That is, if one constructed a series of markups at the aggregate level and performed any number of usual detrending procedures, one typically finds clear cyclical patterns in markups. Moreover, markups are generally found to be *countercyclical* with respect to GDP – that is, periods of lower-than-usual GDP growth tend to be associated with high markups, and vice-versa. An important problem behind any empirical analysis of markups, however, is the appropriate measurement of marginal costs. In our theoretical model, the concept of marginal cost is clear. In an empirical implementation, due to the presence of (sometimes large) fixed costs and **non-constancy** of returns to scale, marginal cost and average total cost are easy to conflate with each other. The theory clearly tells us that it is marginal cost, not average total cost, which is relevant for optimal markups. It is often unclear whether empirical measures of marginal cost are measuring true *marginal* cost or measuring average costs instead.