Dixon, Huw David.

The Role of Imperfect Competition in New Keynesian Economics

pp. 74-108

Dixon, Huw., (2001) Surfing economics: essays for the inquiring economist, Basingstokke: Palgrave

4 The Role of Imperfect Competition in New Keynesian Economics

1 INTRODUCTION

The adjective 'new Keynesian' was introduced in the mid-1980s and refers to a body of work which was published over a period beginning in the mid-1970s. Whilst the term clearly applies to papers written since 1980, it has also been applied retrospectively to some work written before that. Much of this material is gathered together in a two-volume collection of reprinted papers edited by Mankiw and Romer (1991), although the coverage of this volume is somewhat parochial in that only American-based authors are included. There are other more recent surveys, most notably Silvestre (1993) and Dixon and Rankin (1994).[1] In order to understand the phenomenon of new Keynesian macroeconomics, it is essential to set it in an historical context.

2 WHAT'S NEW PUSSYCAT?

The epithet 'new' has been used many times in economics, particularly in recent times. Thus, for example, the 'new' industrial economics; the 'new' trade theory; the 'new' economic geography; are all labels that have come into use since 1980. In these cases the adjective 'new' designates some degree of a break with the 'old', but also some degree of continuity. For example, the 'new' in the new industrial economics literature represents the use of contemporary game theory in the analysis of oligopoly; see for example Vickers (1985), Dixon (1988) and of course the seminal graduate textbook by Tirole (1988). The 'new' in recent international trade literature represents the introduction of imperfect competition into the heart of trade theory.

Of course, new theories and ideas are always coming into being: economists come up with new ideas both from the incentive of theoretical invention, and the need to explain or attempt to understand contemporary economic phenomena. However, the adjective 'new' is introduced when there appears to be a shift in the approach by several economists at around the same time. In effect, a new school of thought or group of people with a common approach comes into being. However, if we look at the history of thought, the epithet 'new' has been used many times. In an academic environment where many people still had an education in Latin, the Latin 'neo' was preferred to the plain English 'new'. For example, the phrase 'neo-classical' refers not only to an architectural style, or the musical idiom of Stravinsky, but also to the integration of perfectly competitive economics both with general equilibrium theory, and with macroeconomics.

In this chapter I shall argue that the fundamental 'new' idea behind new Keynesian models is that of *imperfect competition*. All of the major innovations of the new Keynesian school are made possible or worthwhile only because of imperfect competition. This is the key feature that differentiates the 'new' from the 'old' Keynesians: it differentiates the new from Keynes himself: it differentiates the 'new Keynesian' from the 'new classical' economists. Imperfect competition at its basic level means that agents (firms, households) are not price-takers: they have the power to set prices or wages. Even if all wages and prices are flexible, the presence of imperfect competition in itself means that the economy will be different in a fundamental way from a perfectly competitive economy. Before exploring the story of imperfect competition in the macro context, let us just remind ourselves how special the assumption of perfect competition is, and how it differs from imperfect competition. The fundamental idea can be illustrated within a simple microeconomic framework. The macroeconomic implications will commence after the interlude.

3 IMPERFECT COMPETITION FOR BEGINNERS: A MICROECONOMIC INTERLUDE

There are different ways of defining perfect competition:[2] however, for our purposes in this chapter, we can pick out two important features:

(a) all agents are price-takers;
(b) prices adjust to equate desired supply and demand.

When we say that agents are price-takers, we mean that they treat the 'market price' as given, they believe that they have no ability to influence the market price. Thus, when perfectly competitive firms decide how much output to produce, they treat the price as given and choose the output that equates supply with demand. This decision defines their supply function, which tells us how much they wish to supply at different prices. Similarly with consumers in deciding demand. When we say that prices adjust to equate supply and demand, we mean that the market price is determined (somehow!) at the point where the supply and demand curves intersect at point E in Figure 4.1a, at price P^* and quantity X^*.

One of the most important points to note about the competitive equilibrium is that it is in some senses a socially optimal outcome (in the absence of externalities etc.). In particular, we can say that it *maximises the sum of consumer and producer surplus*,[3] or, more simply, maximises *total surplus*. To see this, note that if we consider the competitive equilibrium in Figure 4.1a, the *producer surplus*, which is best thought of as profits, is given by the area between the horizontal price line $P = P^*$ and the supply curve, represented by the triangle between

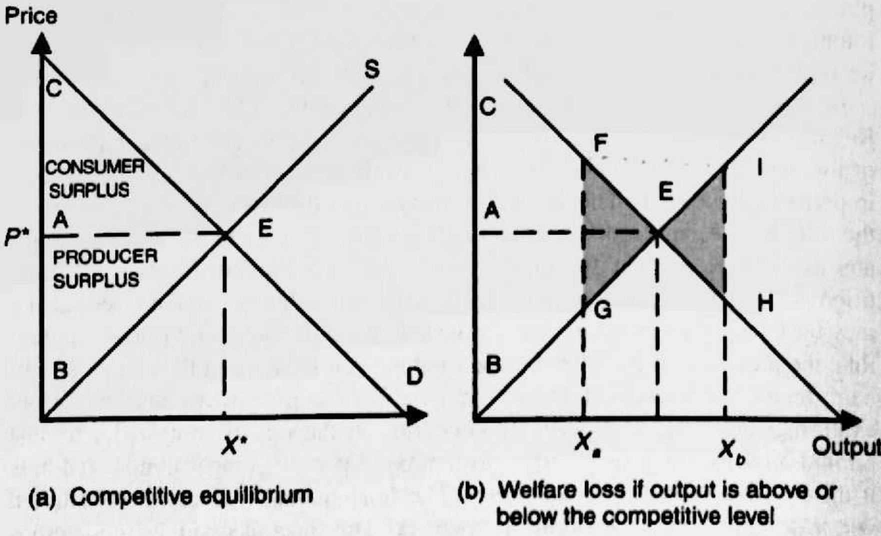
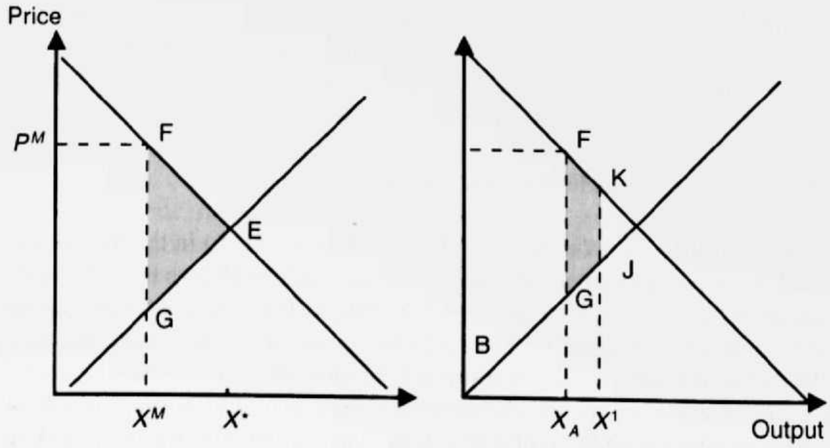


Figure 4.1 The competitive equilibrium maximises total surplus

points ABE. This is because the supply curve is simply the marginal cost curve of the firms supplying the market: thus the additional profit of producing one more unit is the difference between price and marginal cost at the current output. The *consumer surplus* is given by the triangle ACE, the area between the demand curve and the horizontal price line $P = P^*$. *Total surplus* is then the triangle BEC. Note that if the output was below X^*, for example at point X_a as in Figure 4.1b, then total surplus will be less: producer surplus is now given by the unshaded area below the price line, and consumer surplus by the shaded area above the price line. From the point of view of social welfare, the net gain to an additional unit of output is the vertical distance or 'gap' at that output between the demand curve (which represents the marginal value of output) and the marginal cost curve (which represents the marginal cost of output): at X_a this gap is GF. The total loss in surplus when we compare X_a to X^* is the triangle GEF. If output exceeds the competitive level as at point X_b, then this also reduces welfare, since now the marginal cost of output exceeds the marginal value: the loss is given by the triangle EIH.

The lesson of this illustration is that *the competitive equilibrium is in a Pareto-optimal outcome that maximises the sum of producer and consumer surplus (the total surplus). Any deviation from that output, whether it be an increase or a decrease, will tend to reduce the total surplus.*

Now let us consider an imperfectly competitive equilibrium, for example a monopoly. A monopolist will set its price as some mark-up over marginal cost. For example, in Figure 4.2, assume the profit maximising price of the monopolist is P^M, with resultant output X^M. As can be seen if we compare Figures 4.1a and 4.2,



(a) The social cost of monopoly

(b) Increasing output above the monopoly level increases welfare

Figure 4.2 *Welfare loss with imperfect competition*

the monopoly outcome involves a loss in total surplus as compared to the competitive outcome: the net gain in producer surplus to the monopolist is more than offset by the loss of consumer surplus. The total loss is the triangle GEF in Figure 4.2a, which is often called the 'social cost of monopoly'. Thus if we compare the monopoly outcome to the competitive outcome, we observe that in comparison to the competitive outcome, (a) the level of economic activity is lower, and (b) the level of welfare is lower. However the difference does not end there: if for some reason the output is increased beyond X^M, then of course total surplus will increase. For example, if output increases to X^1 in Figure 4.2b, then the gain in total surplus will be the shaded area GFKJ. Thus if we start from an imperfectly competitive equilibrium, then an increase in output will increase welfare.

Hence we can see that there are two fundamental differences between the perfectly competitive equilibrium and the monopolistic equilibrium. First, *the monopolistic equilibrium involves a lower level of welfare than the perfectly competitive equilibrium*. Second, *starting from the monopolistic equilibrium, an increase in output increases welfare, a reduction reduces welfare*. This contrasts with the competitive equilibrium where any deviation of output reduces welfare. Although the above analysis was in terms of an output market, we could think in exactly the same way about a labour market, with P being the real wage, and X the level of employment.

Whilst the analysis of this section has been cast in terms of simple microeconomics, its lessons will carry over into macroeconomics. The extra dimension added in macroeconomics is that the approach is *general equilibrium*: we have to consider equilibrium of all of the markets in the economy, and how they interact.

Now, perhaps it is time to drop the general discussion of the word 'new', and to focus on what we mean by using the word 'Keynesian' after new. Let's take a look back in history to Britain, and more specifically Cambridge in the 1930s.

4 OF KEYNES AND THE KEYNESIANS

When Keynes first wrote *The General Theory* (Keynes, 1936) in the mid-1930s, he in effect gave birth to macroeconomics as a discipline. Before that, the study of large-scale aggregate phenomena such as employment and national income was based on a predominantly microeconomic and partial equilibrium perspective. Even the notion of national income and the measurement of macroeconomic phenomena was not at all developed in a coherent or useful way. Much of Keynes's contribution and that of the earliest macroeconomists was in providing a consistent and useful framework for national income statistics, and founding the accounting conventions that have now become standard.

However, whilst Keynes developed a new theory and new ideas, he was unable to develop a fully integrated framework which was clearly related to the existing approach of 'price theory', or standard supply and demand analysis in either its partial equilibrium version, or its general equilibrium version as developed by Walras. Keynes designated this corpus of theory as 'classical', and he was clear that his theory marked a definite departure from this classical approach. Thus, for example, he starts off *The General Theory* by stating two of the postulates of classical economics, which he defined as:

I *The wage is equal to the marginal product of labour.*
II *The utility of the wage when a given volume of labour is employed is equal to the marginal disutility of that amount of employment.*

(Keynes, 1936, p. 5)

Postulate I states that the labour market outcome is on the (competitive) labour demand curve; postulate II which Keynes rejected states that the labour market outcome is on the labour supply curve. Clearly, the rejection of postulate II introduces the possibility of *involuntary unemployment*, in that if the 'utility of the wage' exceeds the 'marginal disutility of work', then individuals will be willing to work more than they are able to at the prevailing wage. We will discuss this in more detail below.

However, this rejection of classical economics led to a tension within the postwar neoclassical synthesis. On the one hand, in much of microeconomics and subjects such as trade theory, the 'classical' approach largely dominated: agents were assumed to be price-takers; prices (and wages) adjusted so that markets cleared. Agents maximised something subject to some constraint. However, in macroeconomics, this approach was not taken; rather, a series of separate assumptions were made as necessary. For example, Keynes had been willing to assume that consumption was determined by a basic psychological law: he had either not seen it as necessary, or simply did not have the time to tinker with this aspect of his theory so as to show how it was related to the classical case.

The phrase *ad hoc* has often been used to describe this style of macroeconomics: wages were (for example) assumed to be downwardly but upwardly mobile. The most significant development in postwar Keynesian economics was probably the discovery of the Phillips curve (1958), and its integration into macroeconomic models with little or no theoretical underpinning; with the notable exception of Lipsey (1960). Of course, although the phrase *ad hoc* has usually been used perjoratively, there is nothing in principle wrong with *ad hocery* where it is better than the best non-*ad hoc* alternative. Thus Keynes himself was in my opinion quite right to freely develop a model of unemployment which was not fully worked out in the traditional sense when the next best alternative was a model for which unemployment was largely assumed away, at a time during the 1930s when mass unemployment was 'the' major policy issue.

Be that as it may, there was nevertheless a tension between macroeconomics as commonly practised and microeconomics. The success of Keynes's vision of macroeconomics brought to peoples' awareness the need to resolve this tension, and to somehow integrate it with the maximising behaviour which formed the basis of traditional microeconomics. At a more fundamental level, the notion of maximising subject to a constraint is fundamental to the enterprise of explanation by economists.[4]

There have been several different attempts to undertake this synthesis, and in order to understand the distinctive features of new Keynesian thought, it is essential to understand something of these previous attempts at integrating microeconomics and macroeconomics.

5 LITTLE AND LARGE: MICRO AND MACRO

Macroeconomics studies the behaviour of the macroeconomic system, of macroeconomic aggregates. Clearly, there is a relationship between the behaviour of the parts of the system (households, firms, the government) and the behaviour of the aggregates which macroeconomics studies. This relationship is not at all simple. For example, in physics and chemistry it is not always thought of as useful or possible to attempt to derive everything from quantum mechanics. However, in theoretical economics at least, it should be possible to trace through the relationship between the behaviour of the individual agents at the *micro* level in the economy and the behaviour of the economic system at the *macro* level, even if only in a stylised form. The attempt to do this was conceived as a search for the *microfoundations of macroeconomics*. However, this label indicates that the search is one-way: you do not need to consider macroeconomic aspects to get the micro level correct. This is of course not correct: there is a two-way street here, and the behaviour of the economic system at the macroeconomic level can of course influence what the microeconomics needs to be. There is in a sense a need for a *macrofoundation of microeconomics.* Thus when we think of an approach to macroeconomics such as the new Keynesian, we need to think of two levels of theory: the microeconomics of the firm/household/government, and the macroeconomics which corresponds to it. For the theory to be coherent, these two levels need to be consistent.

6 WALRASIAN MICROECONOMICS AND MACROECONOMICS

Léon Walras, a French economist, developed a vision of what we now call a 'general equilibrium system': that is the concept that all markets are linked through the price mechanism, and that in order to balance supply and demand in all markets, it is necessary to have all prices adjust at the same time. Demand for each good in principle depends on the prices of all goods, however indirectly. In his time, Walras was something of a visionary.

There are certainly assumptions which underlie the Walrasian model, which is the general equilibrium version of the standard supply and demand model. At the microeconomic level, it is assumed that agents (firms or households) are price-takers. This means that they believe that they can sell or buy as much as they want at the prevailing market price. This is usually justified by the notion that agents are 'small', too small to affect the market price (although recent work suggests that you can have competitive outcomes with only a few agents). This is a microeconomic assumption. However, in order for this microeconomic assumption to make sense, it is also necessary assume that all markets clear in the sense that prices are in place which equate planned demand with planned supply in all markets. When planned demand equals supply, and only then, can agents on both sides of every market trade as much as they want to at the going prices. This is in effect the *macroeconomic* assumption needed to underpin the microeconomic model.

To see why a global vision is needed, consider what would happen if for some reason the price deviates from that at which demand equals supply. This is depicted in Figure 4.3, where the price P is above the equilibrium price P^*. In this case the desired trades by agents do not match up: agents who are suppliers want to sell more than those demanding the good want to buy. In this framework, we can interpret this market as any market: for example, the 'price' of the 'good' could be the real wage, and the situation of excess supply corresponds to *involuntary* unemployment. In this case the labour that households wish to supply exceeds the quantity demanded by firms. Unemployment in this sense is seen as involuntary in that households are willing to supply more labour than is demanded. However we interpret this market, we can see that the offers by agents to trade (demands and supplies) are inconsistent: whatever happens, all agents cannot realise their planned transactions. Hence in this case to assume that agents behave as if they can buy and sell as much as they want to is hardly satisfactory. This contrasts with the case where the price is continuously at its competitive equilibrium value P^*: in that situation, agents are able to trade as much as they want to.

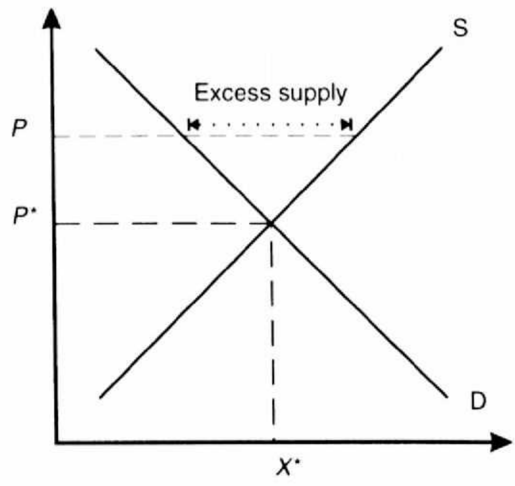


Figure 4.3 *Market clearing in competitive equilibrium*

Before we go on to look at other microeconomic set-ups, let us consider one of the fundamental conceptual problems associated with the Walrasian equilibrium. All agents are assumed to be price-takers, yet prices are assumed to instantaneously adjust to the level where supply equals demand. This is Arrow's paradox, named after Ken Arrow, the Nobel Laureate who pointed this out in his famous article published in 1959. Léon Walras was also aware of this problem, and he was familiar with the operations of the market-makers in the Paris Bourse (the stock exchange which still stands near Les Halles in Paris), who used to set the prices of the various stocks and shares. Léon Walras used the example of this market which he knew well and invented the 'auctioneer' for the whole general equilibrium system. This auctioneer was supposed to determine all prices in the economy (including the factors of production such as labour) so as to equate supply with demand. In a sense, this fictitious auctioneer is a central part of the macroeconomics of the Walrasian system.

7 NON-WALRASIAN MICROECONOMICS: NEO-KEYNESIAN MACROECONOMICS

What happens if prices do not instantaneously clear markets? What happens if people trade at prices other than the competitive or Walrasian prices? Whilst this is a subject that had been thought about by several economists previously, it was not until the 1970s that the subject was examined in full technical detail by the 'neo-Keynesian' economists, most notably Barro and Grossman (1971) and Bénassy (1973, 1975). Let us be clear why the term 'neo-Keynesian' was used in this context.

In the 1960s there was a 'reappraisal' of Keynes, primarily in the works of Clower (1965) and Leijonhufvud (1968). The full story of this reappraisal lies beyond the scope of this chapter, but without doing it full justice I will simply say that two tenets were central to it: first, it was claimed that underlying Keynes's theory was a *disequilibrium* story; secondly, underlying this theory was a coherent if imperfectly articulated microfoundation. The implication of this work was that traditional Keynesian analysis of the 'then' orthodoxy had in effect emasculated Keynes's original insights, and put them in a world of *ad hocery*.

Whatever the details of this phase, the result was the emergence of the neo-Keynesian school in the 1970s. The phrase 'neo-Keynesian' is not a universally accepted term, but it is one used by the main contributor during this phase: namely Jean-Pascal Bénassy. Bénassy studied at Berkeley in California under Gerard Debreu whose *Theory of Value* (1959) ranks as the main classic in the Walrasian tradition (along with John Hicks' *Value and Capital*, 1939). The result was Jean-Pascal Bénassy's thesis (1973), from which came three papers (1975, 1976, 1978) which defined the neo-Keynesian approach. The first contributions in this phase came from Barro and Grossman (1971, 1976), and there is much in common between the work of those authors and Bénassy. There is, however, also a big difference: whereas Bénassy adopted a primarily general equilibrium approach, the work of Barro and Grossman was primarily macroeconomic. This difference is one of perspective: both Barro and Grossman on the one hand, and Bénassy on the other, were trying to develop general equilibrium macromodels, that is microfounded macroeconomics. However, the emphasis was different: whereas Bénassy allowed for many commodities and looked at esoteric issues such as existence, Barro and Grossman adopted the standard aggregation of macroeconomic models and had just three goods (consumption, leisure and money). The neo-Keynesian approach was popularised by Edmond Malinvaud's (1977) book *The Theory of Unemployment Reconsidered*, which made these ideas known and accessible to a general audience; and Muellbauer and Portes (1978) also developed a simple textbook representation which was soon after included in William Branson's (1980) graduate macroeconomics textbook.

What was the essence of the neo-Keynesian school? In the Walrasian framework, all agents are price-takers, and an auctioneer is assumed to ensure that prices instantaneously clear markets, so that demand equals supply in every market. The microeconomics of this had been fully developed in a general equilibrium framework by Debreu and Hicks. The neo-Keynesian school kept the assumption that agents were price-takers, but dropped the assumption that prices adjusted to clear markets. What defines the approach of the neo-Keynesians is the *assumption that wages and prices are fixed*, or at least are *treated as exogenous*.

If the assumption that prices adjust to equate supply and demand is dropped, it then follows that the Walrasian model of firms and households needs to be modified, since it is based on the notion that agents can buy and sell as much as they want. In general this is only true in a competitive equilibrium. The neo-Keynesian school developed the theory of how households and firms would behave if they faced limits on how much they could buy or sell. These models were referred to as *quantity constrained* or *rationing* models (see Levačić and Rebmann, 1982 chapters 16 and 17). However, let us start from the beginning. Looking back at Figure 4.3, we need to consider what will happen in a situation where the price is fixed at a level where supply exceeds demand, as at price *P*. The first step is to establish what trades will take place. In this situation, it is argued that the 'min condition' holds: that is, where supply (S) and demand (D) are different, then the amount actually traded is the minimum (i.e. the smaller) of the two. In Figure 4.3, at price *P*, supply exceeds demand. Thus the min condition tells us that actual trading will be equal to demand D. This is depicted in Figure 4.4. For prices below the competitive price P^*, supply is less than demand (there is excess demand), so that actual trades equal the quantity supplied; for prices above the competitive price, demand is the smaller of the two, so that trades are demand determined. The notion underlying the min condition is simple enough: you cannot force people to trade more or less than they want to – trading is a voluntary activity. Thus, there is no way that the suppliers of a good can be forced to supply more than they want to if demand exceeds supply, and vice versa. In mathematical terms, the min condition can be written as the quantity traded *X* is:

$$X = \min(S, D)$$

Now, recall that in this approach, we treat the price as a exogenous variable, as something fixed. We then trace through the consequences of this. If the price is not equal to the competitive price, then the planned or desired trades of agents are not consistent: they cannot both be realised. Thus one side of the market must have their plans frustrated. This is in contrast to the Walrasian or perfectly

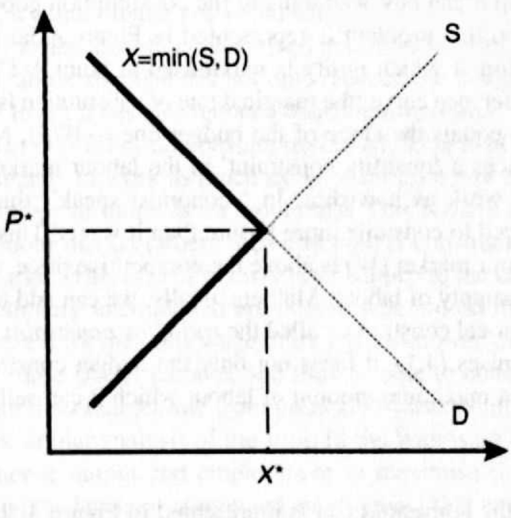


Figure 4.4 ***The min condition***

competitive price at which supply equals demand. Here both sides of the market are able to realise their desired trades.

How will agents respond to finding that they are unable to trade as much as they would like to at the prevailing price level? If someone is unable to buy/sell as much as they want to, then we say that they are *rationed* or *quantity constrained*. A new theory of the firm and household needed to be developed in which agents faced not only the standard budget constraint, but also possible quantity constraints (also called rationing constraints). Let us see very briefly how this theory developed, since it is crucial to understanding how the subsequent new Keynesian developments arose.

Traditional consumer theory assumes that the household maximises a utility function subject to a budget constraint. In a macroeconomic context, the utility function might have utility depending on consumption C and leisure L: the budget constraint might have total expenditure on consumption at price P being less than labour income and profits Π less tax etc. Ignoring taxes for now, and noting that work supplied, N, equals time endowment (set at 1) less leisure, then $N = 1 - L$, and we can write the household's maximisation problem as:

$$\max U(C, L) \tag{4.1}$$
$$\text{s.t. } P \cdot C = W \cdot (1 - L) + \Pi \tag{4.2}$$

The budget line (4.2) is written in such a way that there is no constraint on the amounts of C and L except the 'technological' ones (consumption C has to be non-negative; leisure L has to be less than or equal to the total time endowment 1, and cannot be negative). It is simply a straight line: when there is all play and no work ($L = 1$), consumption cannot exceed the non-labour (profit) income of the household; the slope of the budget line is the real wage, since for every unit of leisure it gives up it can buy W/P units of the consumption good.

The solution to this problem is represented in Figure 4.5a: it is the standard tangency condition at which utility is maximised at point $A(C^*, L^*)$, where the slope of the indifference curve (the marginal rate of substitution between consumption and leisure) equals the slope of the budget line ($-W/P$). Now suppose that the household faces a 'quantity constraint' in the labour market: it is unable to supply as much work as it wishes. In 'economist speak', this means that the household is forced to consume more leisure than it wants. This will occur if the 'price' in the labour market (W) is above the competitive price, and there is a situation of excess supply of labour. Mathematically, we can add to the consumer's problem an additional constraint, called the *rationing constraint*, so that when the household maximises (4.1), it faces not only the budget constraint (4.2), but in addition there is a maximum amount of labour which it can sell, N^R so that:

$$(1 - L) \leq N^R \tag{4.3}$$

The situation of the household can is represented in Figure 4.5b, where the segment of the budget line which involves the household selling more labour than N^R

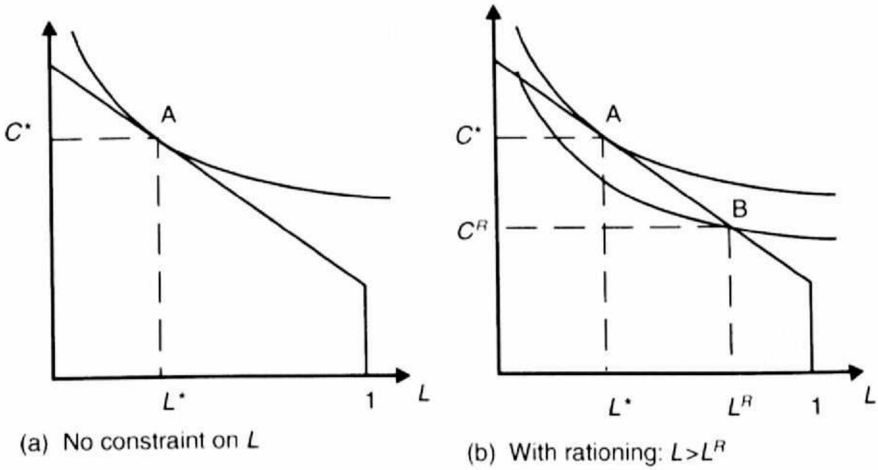


Figure 4.5 *Effective demand*

is now unshaded. The previous optimum (A) given this additional constraint, N^R, is now unfeasible. In effect, rather the consumption/leisure possibilities are represented not by the area between the origin and the budget line (called the 'budget set'), but rather the shaded area in the bottom right-hand corner of the budget set. The maximum utility that the household can attain is now represented by the point along the budget line which involves the household selling up to its constraint, at point B. This must mean a fall in utility from the unconstrained maximum U^* to U^R. This is of course a general result in mathematics: if you impose more constraints, it cannot make you better off!

If we compare the new quantity constrained optimum, we can see that the limit on the amount of labour supplied has not only reduced the labour supply (leisure has risen from L^* to L^R: it has also reduced consumption from C^* to C^R, although there was no direct constraint on consumption itself. This is of course common sense: if we are unable to work as much as we want to, it will mean that we are unable to afford to buy as much as we would like. This is often called a *spillover* effect from the labour market (where the household is constrained) to the output (consumption) market. This is in effect what can happen to the involuntary unemployed. The involuntary unemployed are people who would like to work more than they do at present for the same wage. If they were allowed to work more, then they would have more labour income, and thus be able to consume more. They would move along their budget line from point B towards point A.

We can make a similar analysis of the firm. In the Walrasian analysis, the firm is assumed to choose output and employment to maximise profits: no explicit constraint is put on the levels of output and employment that can be chosen. Suppose that the firm has a standard production function where output y is a function of employment N, $y=f(N)$. Then its profit maximisation problem can be written as one of choosing N (and hence y) to maximise profits:

$$\max P\cdot y - W\cdot N \tag{4.4}$$
$$\text{s.t. } y=f(N) \tag{4.5}$$

The solution to this problem is the standard one that the firm maximises profits by employing labour up to the point where the marginal product of labour (MPL) equals the real wage:

$$\frac{W}{P}=f'(N) \tag{4.6}$$

Thus the demand for labour curve is represented by the MPL curve, which is assumed to be decreasing (due to the diminishing marginal product of labour). From (4.5), we can also determine the desired supply of output by the firm: to the level of employment N^* that solves (4.5) there corresponds the level of output $y^*=f(N^*)$. Clearly, the lower the real wage, the greater the amount of labour the firm will want to employ, and the greater the amount of output it wants to supply. Hence we can write both the demand for labour and the supply of output as functions of the real wage: both are decreasing in W/P:

$$\begin{aligned} y&=y^s\left(\frac{W}{P}\right)\\ N&=N^d\left(\frac{W}{P}\right)\end{aligned} \tag{4.7}$$

Now, suppose that the firm is rationed in the output market, and faces a quantity constraint on the amount of output it can sell. When the firm maximises its profits (4.4), it faces not just the technological constraint (4.5), but also the additional constraint:

$$y\leq y^R \tag{4.8}$$

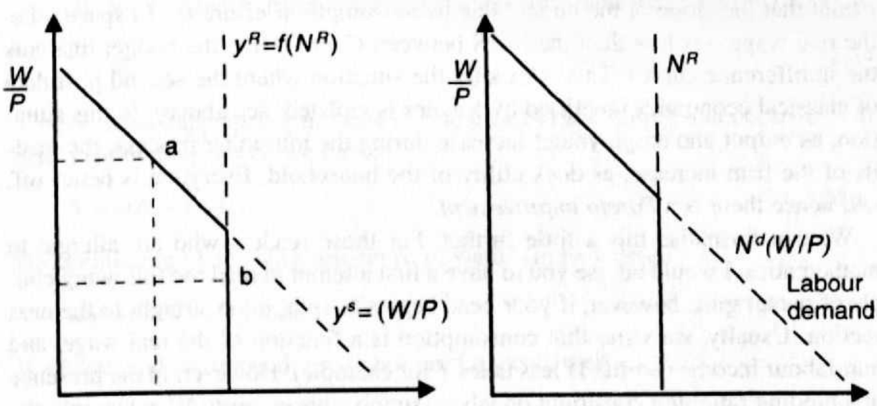
where y^R is the 'ration' on output that can be sold. The solution to maximising (4.4) subject to both (4.5) and (4.8) depends on whether the constraint (4.8) is 'binding': that is, whether the firm wants to sell more or less output than y^R. If the constraint is *non-binding*, then the amount that the firm wishes to sell is less than the quantity constraint y^R. In this case, output and employment by the firm are determined by the output supply function y^s and the labour demand function N^d given by (4.7). However, suppose instead that the firm wishes to sell more than y^R: in this case the firm will want to sell right up to the constraint, so that $y=y^R$. Since the amount the firm would like to sell is determined by the real wage, we can say that actual output y is determined by the following condition:

$$y=\min\left[y^s\left(\frac{W}{P}\right), y^R\right] \tag{4.9}$$

What equation (4.9) tells us is that actual output y is the minimum (that is to say the smaller) of the amount that the firm would like to sell, given by the output supply function $y^s(w/p)$ and the demand constraint y^R. We can represent this in Figure 4.6a: on the vertical axis, we have the real wage, and output is on the horizontal axis; the function $y^s(w/p)$ is downward-sloping (a higher real wage means the firm wants to supply less output): the vertical line $y = y^R$ represents the *demand constraint*. The actual level of output depends on the real wage: if this is high, as at point a, the firm desires to supply less than y^R, so that output is given by the output supply curve y^s. However, at a point like b, where the real wage is low, the firm desires to sell more than the quantity constraint. In this case, we say that the firm is *demand constrained*.

In Figure 4.6b, we can see how the constraint in the *output market* affects the firm's demand in the labour market. The employment level N^R is the exact amount of labour needed to produce y^R: $y^R = f(N^R)$. At high wage levels, such as point a, there is no spillover from the quantity constraint in the output market to the labour market, and employment is determined by the usual labour demand curve.

There is thus an important potential spillover effect operating between rationing constraints in one market (in this example the labour market) and the behaviour of the household in another market (in this example the output or consumption good market). This link is crucial at the macroeconomic level, and forms the foundation for Keynes's theory of *effective demand* and the multiplier. Before we look at this a little more formally, let us consider the argument intuitively. From our example, the optimal consumption decision of households



(a) Demand constraint on output $y=\min(y^R, y^s)$

(b) Effect of output demand constraint on labour demand $N=\min(N^R, N^d)$

Figure 4.6 *The spillover of demand constraints on output onto the firm's demand for labour* depends not only on wages and prices (the factors which along with the profit income determine the position of the budget line represented by (4.5)), but also on the employment decision of firms, which determines the quantity constraint that the household faces. Hence we can get a feedback effect to operate, giving what is usually called the *multiplier*. If someone (e.g. the government) spends more on the output of firms, this will relax the quantity constraint that firms face: they will thus decide to hire more people in the labour market. If firms hire more people in the labour market, then the rationing constraint of the household is relaxed: it is able to supply more labour, which has the effect of increasing wage income and hence consumption. If consumer demand rises, then firms will increase employment; if firms increase employment then consumers will consume more, thereby increasing demand for firm's output.... This is a feedback process, by which the initial first round injection of demand into the economy is magnified. Now, of course, there are various details here that need to be satisfied: both the household and the firm need to be rationed, the firm in the output market, the household in the labour market. When this happens, the economy is said to be in the *Keynesian unemployment* regime (Malinvaud, 1977).

A necessary condition for the firm to be rationed in the output market is that price exceeds marginal cost (MC). To see why, note that the marginal profit from increasing output equals price less marginal cost: if the price is 6 and the MC of an extra unit is 4, the net profit from selling an extra unit at that price is 2 (since $2=6-4$). Similarly, if we look at the household, for the rationing constraint to be binding in the labour market, there must be *involuntary unemployment*, by which we mean that the household is willing to work more than it can at (or even slightly below) the current real wage. One way in which this is often described in the literature is that 'the real wage exceeds the disutility of labour'. By this is meant that the slope of the budget line in consumption/leisure (C, L) space – i.e. the real wage – is less than the MRS between C and L (i.e. the budget line cuts the indifference curve). This is exactly the situation where the second postulate of classical economics identified by Keynes is violated (see above). In this situation, as output and employment increase during the multiplier process, the profits of the firm increase, as does utility of the household. Everyone is better off, and hence there is a *Pareto improvement*.

We can formalise this a little further. For those readers who are allergic to mathematics, I would advise you to have a first attempt to read the following couple of paragraphs: however, if your head begins to spin, jump straight to the next section. Usually, we write that consumption is a function of the real wage, and non-labour income (profits Π less taxes T for example). However, in the presence of a binding rationing constraint on labour supply, this *quantity* N^R enters into the consumption function, representing the spillover from the unemployment in the labour market to consumption:

$$C=C\left(\frac{W}{P}, \frac{\Pi-T}{P}, N^R\right) \tag{4.10}$$

All three derivatives are positive. Let us assume that the firm's technology takes a particular form: one unit of labour is used to produce one unit of output, so that output equals employment.[5] In this case we have $y = N$. We will now introduce the government into the picture: the government purchases g units of output, and pays for it by raising a lump-sum tax on the households T, or running down some asset. Given that, equilibrium in the economy can be written as:

$$N = C(W, \Pi - T, N) + g \tag{4.11}$$

In order to write (4.11), we have assumed that the output price is the numeraire (i.e. we have set $P = 1$).[6] We have also assumed that in the output market $y = y^R = C + g$, and in the labour market, $N = N^R$: using these we have substituted out the rationing constraints and expressed everything in terms of employment N. It is easiest to see what $C(W, \Pi - g, N)$ looks like if we take, for example, Cobb–Douglas preferences $U(C, L) = C^{\alpha}L^{(1-\alpha)}$. The 'unconstrained' consumption function[7] is then:

$$C = \alpha(W + \Pi - T) \tag{4.12}$$

whereas the *constrained* demand when labour supply is limited to N^R is:

$$C = \alpha(W \cdot N + \Pi - T) \tag{4.13}$$

An obvious interpretation of α is the 'marginal propensity to consume', since the household with Cobb–Douglas preferences consumes a proportion α of its income. Now, WN is total labour income, and Π is total profit income. The firm's budget constraint says that total revenue y must be divided between costs (labour is the only factor of production here, so costs equal wage costs $W \cdot N$) and profits Π: that is $y = W \cdot N + \Pi$. Hence, using the firm's budget constraint we can write (4.13) as:

$$C = \alpha(y - T) \tag{4.14}$$

Now, what happens if g increases? Using (4.14) and noting that because of the firm's technology $y = N$, we have:

$$N = \alpha(N - T) + g \tag{4.15}$$

Differentiating (4.15) with respect to g, yields (in two steps):

$$dN = \alpha(dN - dT) + dg$$

If $dT = 0$, and the expenditure is not tax financed, then

$$\frac{dN}{dg} = \frac{1}{1-\alpha} = \frac{1}{1-MPC} > 1 \tag{4.16}$$

This is the classic Keynesian multiplier: the initial stimulus dg is magnified by the feedback process between employment decisions of firms and the consumption decision of households.

Has there been an increase in the welfare of everyone in this process? The simple answer is yes, there must have been. We can tell that from the fact that firms were willing to respond to the increase in the demand by increasing output, and also that the households were willing to supply more labour. Given that there is voluntary trade, any increase in output in this position must lead to an increase in welfare of the household and profits of the firm. In general, how can we tell when this will be the case. This turns out to be quite simple. In order to have a Keynesian multiplier like this, one need to start off from an initial position in which two conditions are satisfied:

(i) There is an excess supply of output: $P > MC$
(ii) There is an excess supply of labour: $W/P >$ disutility of labour.

In order to understand (i), this simply says that the firm would like to sell more at the prevailing price P: a firm will always want to sell more so long as its marginal cost is less than price. Condition (i) simply says that the firm is quantity constrained or rationed in the amount of output it can sell, and hence its profits will increase if it can sell more. A similar observation applies to condition (ii) for households: as in Figure 4.5b, if the household can increase employment (reduce leisure), it can increase utility by moving from B towards A. Note that condition (i) corresponds to the relaxation of Keynes's Classical Postulate I, and (ii) to the relaxation of Classical Postulate II.

There is thus the question: under what assumptions will we be in an initial position where both (i) and (ii) are satisfied? *Here is the crucial link to new Keynesian economics*. In brief, the answer is that imperfectly competitive price/wage setting agents will ensure that (i) and (ii) are satisfied. Turning first to (i): if firms are price-setters, and face a non-perfectly elastic demand curve, then they will set their price as a mark-up over marginal cost: a monopolist, monopolistic firm or oligopolist will set the price above marginal cost as a consequence of profit maximisation. Turning to (ii) there are a variety of different stories. However, if we suppose that the labour market is unionised as in Blanchard and Kiyotaki (1987), then the union will aim to set the real wage above the competitive real wage, and hence to a position where the real wage exceeds the marginal disutility of forgone leisure. The argument is entirely analogous to the firm. Thus, *imperfect competition is crucial in creating the initial condition that households would like to sell more labour, and that firms would like to sell more output at the equilibrium prices*. That is, that *both households and firms are demand constrained*.

However, the mere fact that the initial condition is satisfied is not enough to obtain a Keynesian multiplier: rigid prices are also needed. This is the second key step taken by the new Keynesians: *nominal price rigidity is more likely when there is imperfect competition*. It is to this step that we now turn.

8 NOMINAL RIGIDITY

As we have seen in the previous section, if there is nominal rigidity (fixed wages or prices), then this can give rise to changes in *nominal* demand having *real* output and employment effects. One of the basic insights of new Keynesian economics was to link this idea to that of imperfect competition. This link was made by Michael Parkin (1986), George Akerlof and Janet Yellen (1985a,b), and Greg Mankiw (1985) in what has become known as the 'menu cost' theory.

However, before we go into the details of the menu cost theory, it is useful to briefly review another powerful idea: that of staggered contracts. The original new classical neutrality result of Sargent and Wallace (1976) showed that if prices adjusted instantaneously and agents held rational expectations, then only unanticipated changes in nominal demand could have an effect on real variables such as output and employment, and furthermore that these effects could only last one period. The reason for the transience of the effect was that agents with rational expectations will immediately update their beliefs and expectations in response to the information embodied in the shock. An early response to this was the Fischer (1977) and Taylor (1979) theory of overlapping or 'staggered' contracts: that is they took the basic Sargent and Wallace model, but added the real world assumption that firms/unions[8] do not all adjust prices/wages at once: rather the adjustment of nominal prices is usually spread over the year due to overlapping or staggered contracts. The key result of the staggered wage setting model is that the effects of shocks are no longer limited to the period in which they occur.

For example, suppose that 50 per cent of firms/unions adjust prices in a particular period (there are two groups who have two-period contracts, one group changes each period). Then when a shock occurs only those whose contracts are up for renewal are able to adjust their contracts: the other 50 per cent are still locked into their old contract. Thus the effect of an unanticipated change in nominal demand will last for at least two periods, and there will be what is technically referred to as 'serial correlation' in output: a positive shock will lead to high output for a couple of periods, and a negative shock to low output. However, the early work of Fischer and Taylor, while tracing out the real effects of nominal rigidities, did not provide any explanation of the structure of nominal wage and price setting. What was needed was a theory of why nominal rigidities persist through time.

9 ENVELOPES, MENU COSTS AND NOMINAL RIGIDITY

One of the key new Keynesian ideas was the notion that with price-setting firms, it was possible that nominal prices were more likely to be rigid. The argument is very simple. Suppose that we have a monopolist who sets the price for his good. If he maximises profits, we have the familiar first-order condition that marginal revenue equals marginal cost. Now, of course, at the optimal or profit maximising price, a small change in price will not lead to much change in profits. This is the meaning of the first order condition, that the derivative of profits with respect to price is equal to zero. To see why, let us consider Figure 4.7, which a visual representation of the relationship between price (on the horizontal axis) and profit

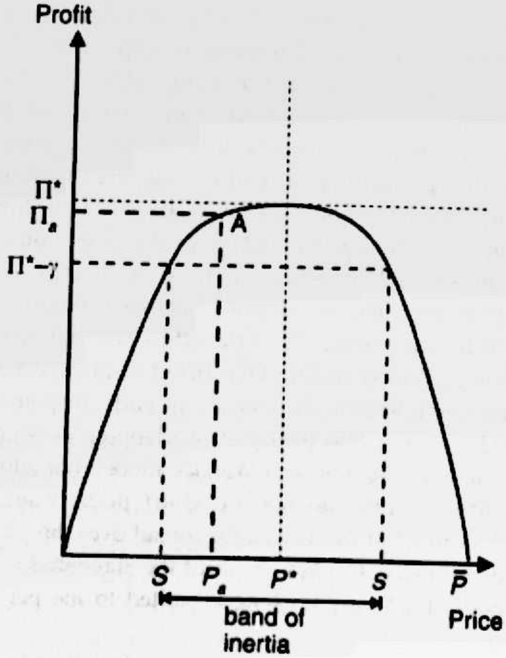


***Figure 4.7** Menu costs and the band of inertia*

(on the vertical). We can see that at a low price (zero, say), profits are zero, and that at a high price $\bar{P}$, the same is true. In between 0 and $\bar{P}$, however, profits are positive, and we have the 'profit hill'. As we set off from a zero price, profits at first increase: we are walking up the slope of the profit hill. However, as we get near to the top, the slope becomes flatter and flatter, until we reach the top. At the top, the hill must be flat: if it had an upward slope, we could not be at the top, since we could get higher by walking up the slope.[9] In mathematical terms, the derivative of profits with respect to price measures the slope: and the first order condition for a maximum states that at the highest level of profits (the summit), profits are flat as we change price (the derivative is zero). Now, this is of course just the intuitive explanation for the first order conditions for a maximum.

However, it has some powerful economic implications, and forms the basis for the 'menu cost' argument for nominal price rigidity. Different authors (Parkin, 1986; Akerlof and Yellen, 1985; and Mankiw, 1985) all saw that if there were some costs to changing prices, then even quite small costs could lead to significant price rigidity. The reason is very simple: suppose that we are near the top of the profit hill, say at point A in Figure 4.7. Furthermore, suppose that it costs a certain amount γ to change the price. In this case, we will only incur the cost of changing price if the benefit we derive in terms of extra profits is larger than γ.

Now as we can see, if we are at A, the benefit is less than γ: on the vertical axis the maximum profit is Π^*, and the profits at point A are Π_a, which is greater than $\Pi^* - \gamma$. Indeed, the fact that the hill is flat near to the top means that although we are not far from the top in terms of profits (the vertical height), the price might be quite far away from the optimal price P^* (the horizontal distance $P_a - P^*$). There is in fact a 'band of inertia' around the optimal price, representing all of those prices like P_a where the cost of changing price outweighs the benefits: this band is represented in Figure 4.7 by the range of prices on the horizontal axis between s and S; the reason for this notation will become clearer below, when we discuss (s, S) rules.

Whilst this all sounds as if it has more to do with geography than macroeconomics, that is not so. Suppose that there is a change in demand, for example the demand curve shifts. In terms of Figure 4.7, the mountain would move, or at least the profit hill would shift in some way: suppose that the whole thing might move to the right. Now rather than drawing the situation 'before' and 'after' the mountain moved, we can simply reinterpret Figure 4.7: P_a is the old optimal price before the mountain moved, and P^* is the new optimal price, and the profit hill drawn is the 'after' one. We can now see that if there is not a big move in the mountain (i.e. demand does not change too much), then the gain from adjusting our price (increasing profits from Π_a to Π^* will not be very much, since the old optimal price P_a is still near to the top of the hill, and the slope of the hill is flat. This mathematical result is known as the envelope theorem.[10]

What conclusion can we draw from this analysis? If a price-setting monopolist has some costs of adjusting price, then if there is a small change in demand (the mountain only moves a little bit), then he will not change price. Even quite small menu costs can lead to significant price rigidity. Indeed, it has long been observed that there is significant price rigidity in imperfectly competitive markets, and this is a possible explanation. However, whilst the menu costs might be quite small, as we have seen the welfare benefits of an increase in demand can be large. We know that a monopolist will mark up price over marginal cost: the more inelastic demand, the higher the gap between price and marginal cost. In terms of social welfare, the market price of the output exceeds the social costs of production (which will be equal to marginal cost if the factor markets are perfectly competitive). The increase in welfare is shown in Figure 4.8.

The demand curve Da is the initial demand curve corresponding to the optimal price P_a (you can link together Figure 4.7 and Figure 4.8): demand increases to Db, and the optimal price 'after' is P^*. However, suppose that the change in demand is small,[11] so that the potential gain in profits Π_a to Π^* from changing price from P_a to P^* is less than γ, so that the price remains fixed at P_a and output will have increased from X_a to X'_a. In this case there is a gain in total welfare: for each extra unit of output produced, the marginal value of this output to consumers (represented by the market price P_a) exceeds the social cost of production (represented by the marginal cost MC, here given). Note that the firm will of course earn more profits: the marginal profit on each extra unit sold is the

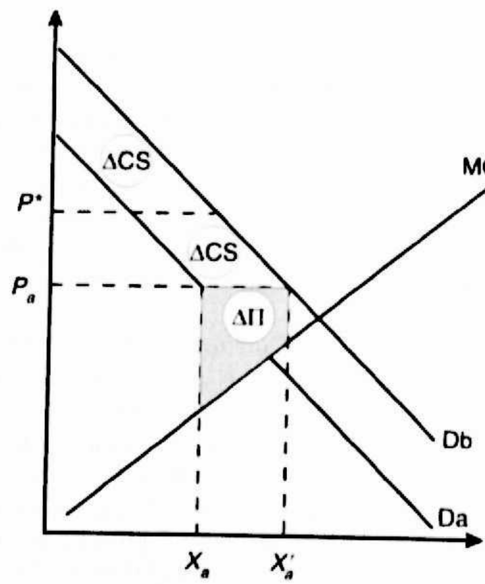


Figure 4.8 *Pareto improvement with a rigid nominal price*

difference between MC and price, the total increase in profits being represented by the shaded area ΔΠ. Furthermore, there is an increase in consumer surplus, the shaded area ΔCS (the whole area between the demand curves above the horizontal line P_a. Hence, if demand increases *and the price is rigid due to menu costs*, then there can be a Pareto improvement – consumers and the owners of firms are both made better off. Of course, the firm's shareholders would be even better off if there were no menu costs and the firm could raise its price to P^*: but the point remains that nominal rigidity of prices means that an increase in nominal demand can lead to a Pareto improvement whereby everyone is better off. This is obviously a very Keynesian result: if we interpret the increase in demand as being due to a tax cut or an increase in the nominal money supply, or a money financed increase in government expenditure, then when there are menu costs this sort of policy can lead to 'Keynesian' multiplier effects (since prices are rigid), and also 'Keynesian' welfare effects, since there is a Pareto improvement.

Now, all of this argument rests on the importance or size of the so called 'menu costs'. If menu costs are extremely small, then they will not cause much nominal rigidity. In terms of the band of inertia in Figure 4.7, the range (s, S) will be quite small. This means the ability of the government to take advantage of the rigidity is quite limited. There are three explanations of menu costs. Firstly, there is the literal cost of printing new price lists and informing customers of a price change. For a restaurant, this can be quite trivial. However, for some types of organisation this alone can be quite significant: for example, large banks and other institutions that provide variable interest home loans may have millions of customers. The cost of sending a letter to each one of these to inform them of a change in the loan interest rate (which they are required to do by law) is in itself very large (printing and postage costs). However, with a few specific exceptions, not many people think that these narrowly defined menu costs are what count. Rather there are two other types of argument which are really different ways of looking at the same thing.

First, there is the argument of bounded rationality (this was the original argument of Akerlof and Yellen, 1985): people do not really maximise, they adopt 'near rationality' in the sense of taking actions which tend to get them close to the optimum. Why should they behave in this way? This brings us to the second argument, which is that there are costs to decisions; we need to gather information, process it, decide what to do, and then implement our decision. These 'decision costs' mean that firms often adopt 'rules of thumb', guidelines for such activities as price-setting that seem to work. Indeed, one can put together the two arguments and say that bounded rationality is really full rationality but with decision costs taken into account.[12] Indeed, there has been a long tradition of literature which has argued that firms adopt simple 'rules of thumb' in their pricing decisions (see for example Hall and Hitch, 1939, and for a more recent and explicitly macroeconomic approach, see Naish, 1993). Certainly, there is strong empirical evidence that for variety of reasons some firms do not change their prices all that often (indeed, it is rare to find restaurants that vary prices frequently, despite the small menu costs).

However, even if we take the menu cost theory at the individual firm level, what are the macroeconomic implications? Caplin and Spulber (1987) came up with an interesting argument which showed that even if menu costs lead to price rigidity at the *firm* level, they might not lead to any significant rigidity at the *macroeconomic* level. In order to understand the argument, we need to look a bit more at the so-called (s, S) rule. The argument that we have advanced about menu costs was static: it looked at the effect of a one-off change in demand. In practice, of course, firms need to consider what is going to happen in the future when they set their prices now. If the general level of nominal prices in the economy is rising (there is background inflation), then setting a lower price now will lead to having to change prices sooner (and hence incur menu costs sooner) than if a higher price were set now. In practice there is a trade off between setting a price to optimise *current* profits, and the need to take into account future profits. This is a complex problem in dynamic optimisation which I will not explain here. However, some clever chaps have solved this sort of problem, and the solution is that firms adopt a (s, S) rule, which is really quite simple.

First, consider the optimal price in the absence of any menu costs at each instant t of time: $P^*(t)$. This is the price that equates current marginal revenue with MC. The optimal pricing rule[13] takes the following form. There is a lower barrier $P(t)^* - s$, and an upper barrier $P^*(t) + S$: these two barriers together define a band of inertia around the price $P^*(t)$ (and hence the two parameters (s, S)

define this sort of rule). If the price at any time is within the band of inertia, then the (s, S) indicates that it is optimal to leave the price where it is. However, if the price moves outside the boundary, then it should be adjusted, with the exact rule for setting the new price depending on the expected behaviour of future prices. We depict the sort of pricing behaviour by an individual firm when there is a constant background inflation, so that the 'optimal' price $P^*(t)$ follows a smooth upward trend equal to the rate of inflation, as in Figure 4.9.

The optimal pricing rule involves the following behaviour: the price is kept constant until the actual price hits the bottom barrier; at that point the firm will raise the price *above* the $P^*(t)$, and then hold it constant.[14] So, in Figure 4.9, we can see that the optimal behaviour involves the firm keeping the price constant most of the time, and having a periodic revision of the price to keep in line with inflation (depicted at time t_0 and t_1). This is realistic: we often observe firms which seem[15] to publish new price lists at regular intervals (every quarter or every year). Now Caplin and Spulber argued that this type of behaviour at the micro level is perfectly consistent with complete price flexibility at the aggregate macro level.

The argument is really quite simple. Imagine that we are in Ancient Rome,[16] and that there are ten months in the year. There are many firms, each with menu costs, following a (s, S) rule. The background inflation rate is 10 per cent per annum: each firm finds it optimal to change its price once per year. Thus, in terms of Figure 4.9, each individual firm keeps its price constant for 9 months of the year, and on the 10th it changes its price by 10 per cent. Now (and here is the

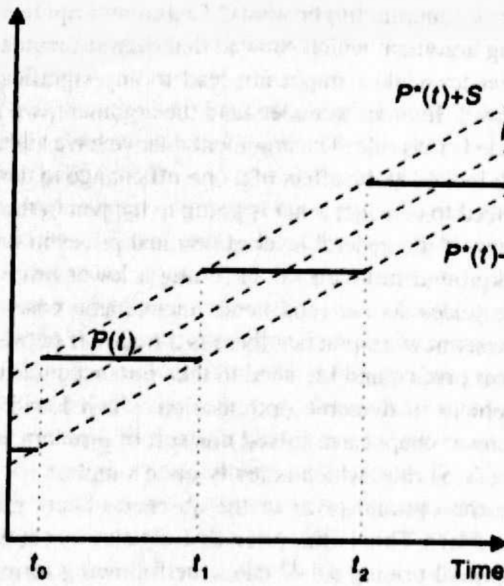


***Figure 4.9** Optimal pricing under an (s, S) rule*

interesting bit), if the critical month at which each firm changes its price are evenly spread over the year (1/10 each month), then the monthly inflation rate will be 1 per cent. To see why, in any one month 9/10 of prices are fixed, and 1/10 rise by 10 per cent: the average is therefore 1 per cent [(0 per cent × 0.9) + (10 per cent × 0.1) = 1]. Thus, each month we can see prices *on average* rising by 1 per cent: and over the 10 months there is a cumulative increase of 10 per cent.[17] Thus, although there is a nominal rigidity at the individual firm level, the aggregate price time series is perfectly smooth. Caplin and Spulber's argument rests on the notion that the incidence of price adjustment through time is even: each month sees an equal proportion of firms changing price. In fact they made a lot of very special assumptions to ensure that this was the case, and it is not a general result. Alan Sutherland (1995) analysed this in somewhat more detail, and found that 'clustering' was a more common phenomenon: firms would tend to arrange to change their prices together. In that case the aggregate price index would not be so smooth. In practice, we often observe such 'clustering': lots of prices change just after Christmas, and so on. Hence microeconomic price rigidity can lead to macroeconomic price rigidity.

One of the main empirical tests of the menu cost theory suggested by Ball, Mankiw and Romer (1988) was that there would be a relationship between inflation and the responsiveness of output to *nominal* demand shocks. In a high inflation country, the frequency with which firms adjust their prices will be higher. Hence any potential nominal rigidity will be less persistent, so that if there is a nominal demand shock (e.g. an unanticipated change in nominal national income), then the possibility of it translating into a *real* output change is less. The main prediction of the menu cost theory is therefore that the translation of nominal demand shocks into real output changes will be less in high inflation countries. They found some empirical evidence for this relationship looking at a large cross section of countries over a couple of decades.

10 IMPERFECT COMPETITION AND THE MULTIPLIER WITH FLEXIBLE PRICES

Imperfect competition plays a crucial role in the theory of new Keynesian macroeconomics. As we saw in the analysis of menu costs, it provides the basis of a theory of *nominal* price rigidity. However, it also provides the foundation of a theory of *real* rigidity: imperfect competition is an alternative equilibrium concept to the Walrasian one where supply equals demand. Imperfect competition provides an explanation of *how* prices are set by optimising agents rather than by fictitious auctioneers. However, the importance goes further than that, since the imperfectly competitive equilibrium may well be one where price exceeds marginal cost, and if the labour market is unionised, one where there might be involuntary unemployment. In short, the imperfectly competitive equilibrium might be 'Keynesian' in some sense. It is this possibility that I explore next. What does an imperfectly competitive economy without menu costs look like: is it possible to get something that is Keynesian even when prices and wages are perfectly flexible? I will look at this in two stages: first, I will examine an economy in which the labour market is competitive, and the only imperfection is that the product market is imperfectly competitive; secondly, I will look at an economy in which the labour market is not perfectly competitive.

10.1 Imperfect Competition in the Product Market

A few papers have looked at the effect of imperfect competition on the size of the fiscal multiplier (Dixon, 1987c; Mankiw, 1988; Startz, 1989; Marris, 1991; Dixon and Lawler, 1996). Three of the authors have made the claim that in some sense imperfect competition makes the economy Keynesian, and in particular that the traditional Keynesian multiplier $[1/(1-MPC)]$ can in some sense be said to arise in an imperfectly competitive economy. Let us look a bit more closely at this claim.

In order to keep things ultra simple, let us consider an economy in which labour is the only factor of production, and the marginal product of labour is equal to unity: output equals employment: $y=N$. There are two goods, leisure $L=1-N$, and consumption C. Households have the utility function

$$U(C,L)=C^{\alpha}L^{(1-\alpha)}$$

subject to the standard budget constraint we discussed earlier:

$$C=W\cdot(1-L)+\Pi-T \tag{4.17}$$

Again, we are treating the output good as the numeraire. We can set up the Lagrangean for this problem as follows:

$$\mathfrak{I}=C^{\alpha}L^{(1-\alpha)}+\lambda[W\cdot(1-L)+\Pi-T]$$

The first order conditions are then:

$$\frac{\partial\mathfrak{I}}{\partial C}=\alpha C^{\alpha-1}L^{1-\alpha}-\lambda=0$$

$$\frac{\partial\mathfrak{I}}{\partial L}=(1-\alpha)C^{\alpha}L^{-\alpha}-\lambda W=0$$

From the first equation we have[18] $\lambda=\alpha U/C$, and from the second equation we have $\lambda=(1-\alpha)U/WL$. Combining these two, we have the equation:

$$\frac{C}{L}=\frac{\alpha}{1-\alpha}\cdot W \tag{4.18}$$

We can represent this graphically in Figure 4.10. The household wants to consume consumption and leisure in fixed proportions, so that C/L is determined by its preferences (α) and the real wage (W) it faces. Thus, for any given value of W, the desired ratio C/L can be represented as a ray from the origin, with slope

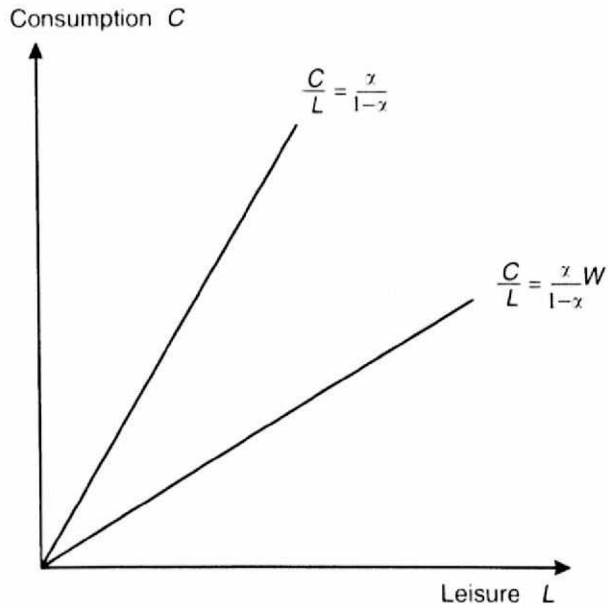


Figure 4.10 *Income expansion paths in consumption–leisure space with Cobb–Douglas preferences*

$(1-\chi)/\chi W$. In Figure 4.10, we have depicted two such rays,[19] one corresponding to a wage of 1 and the other a wage $W < 1$. As the real wage falls, the household responds by consuming less and enjoying more leisure, since the rewards to work are less.

To find the actual level of consumption, you simply substitute equation (11) into the budget constraint (4.17), so that:

$$C = \chi(W + \Pi - T) \tag{4.19}$$

Now comes the relevance of imperfect competition. Firms produce one unit of output with one unit of labour, so that marginal cost (MC) is W. Let us simply assume that there is some imperfect competition, so that the typical firm[20] is able to mark up the price over marginal cost: in particular, the so called 'price–cost margin':[21]

$$\mu = \frac{P-W}{P} = 1 - W \tag{4.20}$$

Since we are treating the output price as the numeraire, this has a very simple form. The meaning of (4.20) is really quite simple: the more monopoly power the firm has, the more it is able to mark up price over marginal cost, and hence the larger is $P - W$ relative to P (i.e. the larger is $1 - W$), and μ is large. In the case of perfect competition, the firm sets a price equal to marginal cost, so that $W = 1$, and $\mu = 0$: there is no monopoly power. Note that μ lies in the range 0 to 1: even if wages are 0, μ is still only 1. In fact, it is more useful to invert (4.20), and to express W as a function of μ, so that:

$$W = 1 - \mu \tag{4.21}$$

The real wage that the household receives is decreasing in the market power of the firm. This makes sense: each unit of labour produces one unit of output, and this unit of output is divided so that μ goes in profits[22] and $1-\mu$ in wages. Total profits are in real terms μN and total real wages are $N(1-\mu)$. If we consider equation (4.21), we can see that with imperfect competition in the product market the first postulate of classical economics is broken: unless there is perfect competition ($\mu = 0$), then the real wage is strictly less than the marginal product of labour (which equals one in our example).

So, what has all this got to do with macroeconomics? Well, quite a lot, because with these simple equations we can find out the effect of imperfect competition on the fiscal multiplier. If we add government expenditure to the consumers expenditure, and substitute for wages and profits in terms of μ and we have the income–expenditure equilibrium in the output market:

$$y = C + g = \alpha(W + \Pi - T) + g$$

Since $y = N$, $W = 1 - \mu$ and $\Pi = \mu N$, also assuming a balanced budget ($g = T$), this becomes:

$$N = \alpha[(1-\mu) + \mu N - T] + g$$

so that:

$$N = \frac{\alpha(1-\mu) + g(1-\alpha)}{1 - \alpha\mu} \tag{4.22}$$

This is the exact solution for equilibrium output and employment. As we illustrate in Table 4.1 below, as μ increases (firms have more market power), the level of output and employment decrease.[23] This is a standard result, which in no way depends on the functional form we have chosen. In order to obtain the multiplier, we differentiate (4.22) with respect to g:

$$\frac{dN}{dg} = \frac{1-\alpha}{1-\alpha\mu} \leqslant 1 \tag{4.23}$$

Equation (4.23) is very interesting: it shows that there is a direct link between the market power of firms μ and the size of the expenditure multiplier. Note first that the multiplier must be less than 1: even if μ takes its largest possible value of 1, the multiplier is just equal to unity. However, for all practical values of μ, the multiplier will be less than 1. This means that there is some crowding out of consumption, which is not surprising given that the increase in expenditure is financed by tax. Second, note that in a Walrasian world with perfect competition,

Table 4.1 *The relationship between the multiplier, output and the degree of imperfect competition*

μ	0.1	0.2	0.3	0.4	0.5
N	0.837	0.821	0.803	0.780	0.750
$\frac{dN}{dg}$	0.217	0.238	0.263	0.294	0.333

Note: Obtained from equations (4.22) and (4.23) setting $\alpha=0.8$ and $g=0.25$.

$\mu=0$, and the multiplier is:

$$\left.\frac{dN}{dg}\right|_{\mu=0}=1-\alpha \tag{4.24}$$

Now what happens as we increase μ? From (4.23), it is clear that *an increase in μ increases the multiplier*: an increase in imperfect competition leads to an increase in the value of the multiplier. Let us take an example: suppose that $\alpha=0.8$ (a very plausible value if we interpret α as the marginal propensity to consume). In Table 4.1 we give the value of the multiplier for different values of μ. For reference, we also compute the equilibrium output and employment level N, given $g=0.25$. If we compare the Walrasian value (0.2) with the plausible empirical value for μ of 0.3, we can see that the multiplier is 19 per cent larger under imperfect competition. This means that the amount of expenditure necessary to yield a given increase in employment is smaller.

Whilst we can see the mathematics quite clearly, what is the intuitive reason behind this result? All three authors (Dixon, Mankiw and Startz) provide the same explanation, in terms of the *profit multiplier*, which is really quite simple to understand. Suppose that the government increases expenditure by an amount dg. Now, this will be received by the firms as income: they will pass some of the income to households in the form of profits. The initial increase in output associated with the increased expenditure is $dN=dg$: the extra profits resulting from this are then $\mu\cdot dg$, which will then appear in the household's budget constraint in the form of profits. The household will (from equation (4.19)), decide to spend a proportion α of this, thus causing an additional increase in output of $\alpha\mu dg$, and so on.[24] If there is perfect competition and no profits, then there can be no profit multiplier: but with more imperfect competition and a larger mark-up, this effect will be more powerful. Whilst we have looked at the impact of imperfect competition of the government expenditure, it will also apply to other real shocks, such as productivity and real exchange rate shocks.

We can show the effect of imperfect competition on the multiplier diagrammaticaly in Figure 4.11. The vertical and horizontal axes are consumption and leisure as before, and the income expansion paths correspond to those in Figure 4.10. The new feature is to include the production possibility frontier for the case

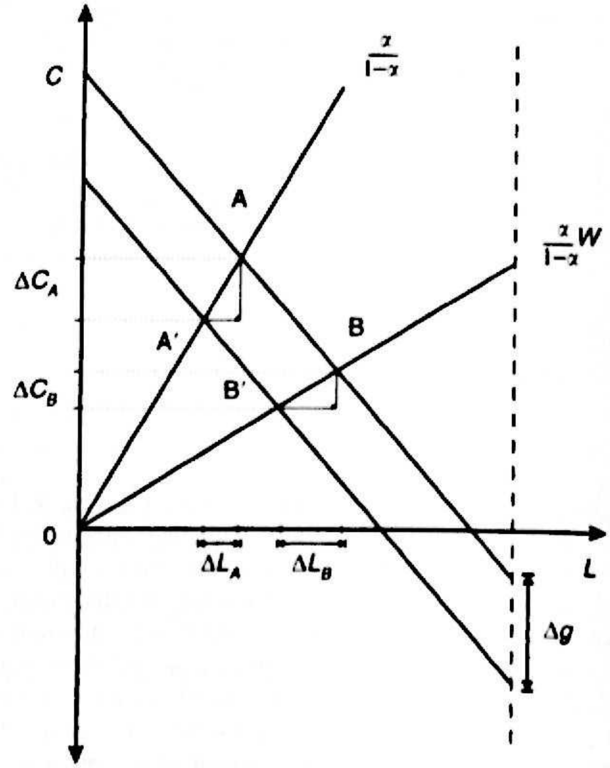


Figure 4.11 *Imperfect competition and the multiplier*

where there is government expenditure. There is one unit of time allocated to the household: it can spread this between work to produce output and leisure:

$$y+L=1 \tag{4.25}$$

We can think of this as the production possibility frontier (PPF) for the economy. However, since output is divided between C and g, we can rewrite (4.25) as:

$$C=1-L-g \tag{4.26}$$

This is represented by the downward-sloping 45° line in Figure 4.11. Clearly, if $L=0$ (the household works all of the time), then $C=1-g$; this is the intercept term for the PPF on the consumption axis. If $L=1$ (the household does not work at all), then consumption should be equal to $-g$, a negative number: this is why we have allowed for negative consumption in Figure 4.11 (this makes for graphical clarity – the household would of course never choose $1-L=N\leqslant g$).

An increase in g means that the PPF in (C,L) space shifts downwards by the size of the increase in g, that is Δg. Now the initial equilibrium for the economy will occur where the relevant income expansion path intersects the production possibility frontier: in the Walrasian case at A, or for the case with imperfect competition at B. After the increase in g, the new equilibria will be A' and B' respectively. Clearly, in both of these cases, the level of consumption has been reduced in response to the increase in government expenditure (there is crowding out): however, the reduction is less than the increase in government expenditure (there is less than 100 per cent crowding out). In the Walrasian case, the reduction in consumption is ΔC_w, and in the imperfectly competitive case ΔC_μ. Clearly, since the slope of the imperfectly competitive IEP is less than the Walrasian IEP, it follows that the degree of crowding out is less, since:

$$\Delta C_w > \Delta C_\mu$$

Thus, the reason that the multiplier is greater in the imperfectly competitive case is that there is less crowding out.

As we can see from the above analysis, there is an important issue as to whether the multiplier is Keynesian or not: in Dixon (1987c), I called the multiplier 'Walrasian', since the mechanism by which output increases is that households are made worse off (since leisure is a normal good, if the labour supply increases, then the household must be worse off if the real wage is unchanged). Others (Mankiw, 1988; Startz, 1989) have interpreted such effects as Keynesian. However, whatever interpretation one has, the clear message is that imperfect competition matters here.

10.2 Imperfect Competition in the Labour Market

Whilst imperfect competition in the output market alone can give rise to some Keynesian effects, it cannot explain *involuntary unemployment*. If the labour market is perfectly competitive, then real wages will be such that households will be able to supply all of the labour they wish. Whilst there may be *underemployment* in the labour market (in the sense that the level of employment is lower than in the Walrasian equilibrium), any unemployment is voluntary.

Let us look a little bit more closely at the nature of underemployment, and consider again the model of the previous section. Equation (4.21) can be interpreted as the *demand curve* for labour: it states that the real wage W equals the marginal product of labour (which was assumed to equal 1) times $(1-\mu)$. The usual demand for labour curve is of course assumed to be downward-sloping, because it is usual to assume a diminishing marginal product of labour. However, whether the marginal product of labour is constant or decreasing does not alter the argument. Suppose we depict a labour supply curve, and suppose that the labour supply depends only upon the real wage as depicted in Figure 4.12. In this case as μ increases, the labour demand curve shifts downwards: and hence the equilibrium level of employment decreases. The fact that employment is below its Walrasian level when $\mu > 0$ is defined as *underemployment*.

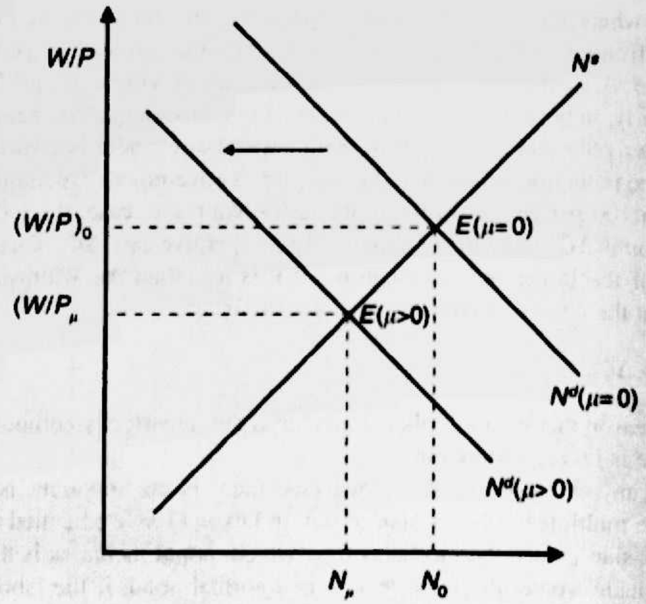


Figure 4.12 Equilibrium employment and imperfect competition in the product market

However, involuntary unemployment arises only if the household is off its labour supply curve. Imperfect competition is a way of explaining why this might be the case. The simplest idea is to imagine that the household/union acts as a 'monopolist' in its supply of labour, see for example Blanchard and Kiyotaki (1987), Dixon (1987c). It is able to restrict the supply of labour in order to increase the real wage (in effect it acts as union). For example, suppose that the union likes employment and real wages: that is, it has a union utility function defined over real wages and employment $V(W,N)$. Assuming that these have the usual properties of utility functions, we can represent them by downward-sloping indifference curves that are convex to the origin. Suppose that the technology of firms displays the usual diminishing marginal product of labour, so that we have the standard downward-sloping demand for labour. The utility-maximising union will choose the real wage and employment level so that the indifference curve is tangential to the labour demand curve, as at V^* in Figure 4.13. We have also drawn in the usual labour supply curve: the union will choose to restrict the level of unemployment to a position as represented in the figure at U, at which the marginal disutility of labour is less than the real wage: that is, there is involuntary (or 'union-voluntary') unemployment, represented by the horizontal distance from U to the labour supply curve. The equilibrium with perfect competition and no union is represented by N_0; the equilibrium level of employment and the real wage with imperfect competition only in the output market is represented by

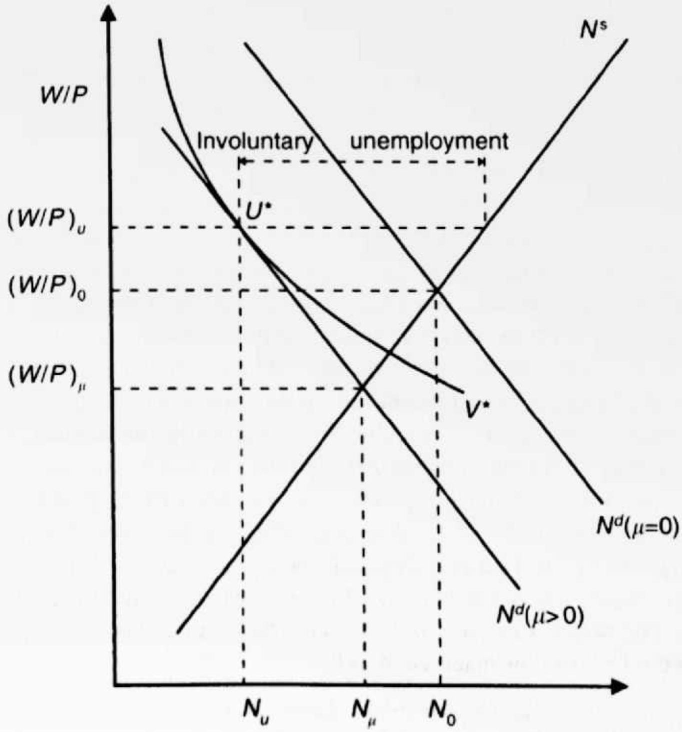


Figure 4.13 *Involuntary unemployment with a unionised labour market*

(W_μ, N_μ); the unionised equilibrium with the imperfectly competitive output market is (W_u, N_u). Clearly, $N_0 > N_\mu > N_u$; furthermore $W_0 > W_\mu$ and $W_u > W_\mu$; the relationship between W_0 and W_u is in general ambiguous, although we have depicted the case where the unionised wage with an imperfectly competitive output market exceeds the Walrasian wage.

11 THE CAMBRIDGE SOAP: WHAT MIGHT HAVE BEEN

In this chapter I have argued that imperfect competition is the key aspect of new Keynesian economics. It is interesting to look back and ask why did it not feature in Keynes's writing, or indeed in subsequent 'Keynesian' writing. On the former, Robin Marris (1991) has written definitively on the subject of Keynes and imperfect competition, and there is little doubt that Keynes rarely thought about imperfect competition. However, this is in a sense very surprising since at the same time as Keynes was developing his macroeconomic theory, the theory of imperfect competition was being developed at the same university by Joan Robinson and Richard Kahn. The explanation appears to have to do with social mores of

Cambridge in the 1930s, and personal tensions between the three. Unlike Keynes, Kahn and Robinson were very much heterosexual. Indeed they had an affair, with Keynes more than once entering Kahn's room to find them both on the floor. As Keynes observed: 'though I expect the conversation was only on the pure theory of monopoly' (Skidelsky, 1992, esp. pp. 448, 495, 536). The ensuing discussion no doubt included the topic of imperfect competition, and the resultant creation was Robinson's (1933) *The Economics of Imperfect Competition*.

Although Joan Robinson did indeed discuss the *General Theory* with Keynes, the link between imperfect competition and the ideas of the *General Theory* was never made.[25] It had been Richard Kahn (1931) who first thought of the multiplier and who helped Robinson (1933) develop her own theory of imperfect competition. Had Kahn and Keynes been able to work together, or Keynes and Robinson, the *General Theory* might have been very different. Another 'K' is of course Kalecki, a much more sensible person who ended up at Oxford. He certainly made the link between imperfect competition and Keynesian economics. However, the idea was buried in a review of Keynes written in Polish in 1936 which was not translated into English until 1982. Kalecki never developed the idea in English, nor in Polish so far as I know. Thus, although there are many 'might have beens', it is clear that in Keynes's writing, imperfect competition played no role, and it was really only with the new Keynesians that the idea was pushed to the forefront of macroeconomics.[26]

12 CONCLUSION

In this chapter, I have explored the key insights of new Keynesian economics as I see them. It is of course something of a presumption to batch together a range of individual people and denote them as 'new Keynesian'. Some individuals have certainly called themselves 'new Keynesian' (most obviously Greg Mankiw and David Romer): others have acquiesced in being so called.[27] However, there are certainly some common themes that seem to be shared in the ideas that we have explored in this chapter, and I will draw these together in this conclusion.

In a perfectly competitive or 'Walrasian' world, the price mechanism ensures that the economy is Pareto-optimal. Even if there are fluctuations in output (due, for example, to changes in technology and so on, as stressed by Real Business Cycle Theory), these fluctuations are optimal: just as a farmer 'makes hay whilst the sun shines', a prudent firm will try to produce more output in periods which are favourable to production. Deviations of output from the perfectly competitive equilibrium have no first order effects on welfare, and increases in output above the equilibrium will if anything tend to reduce the level of welfare.

However, an imperfectly competitive world is inherently non-Pareto-optimal: in such a world, fluctuations in output can have positive (negative) first-order effects on welfare. If there is equilibrium involuntary (union voluntary) unemployment, then an increase in output and employment can increase profits and the welfare of workers. There can be a Pareto improvement, with everyone better off. The microeconomics of the consumer and the firm with arbitrary fixed wages and prices was developed and perfected in the 1970s, and this was well-understood. The key contribution of new Keynesian economics has been to use imperfect competition as a foundation for an equilibrium in which firms and households both want to sell more, and also as a theory of nominal rigidity.

I started off this chapter by looking at the word 'new' in economics: how it applied to such areas as the new industrial economics and the new trade theory in the 1980s. I will conclude with the observation that in all of these fields, much of the 'newness' has arisen from the introduction of imperfectly competitive models into what were before either *ad hoc* or Walrasian approaches. In this sense, the new Keynesian macroeconomics is simply one aspect of the increasing recognition of economists of the importance of imperfect competition in explaining the economic world in which we live.

Notes

1 The latter is included in a collection entitled *The New Macroeconomics*, edited by Dixon and Rankin.
2 In macroeconomics, whilst we talk of perfect competition, the term 'Walrasian equilibrium' is often used instead of 'perfectly competitive equilibrium', in deference to the work of Léon Walras (of which more anon).
3 We do not wish to enter into the details of different measures of social welfare here: any good intermediate micro text will have a lot to say on alternative measures of welfare.
4 See Dixon (1990) for a detailed exploration of this theme.
5 This may seem a rather odd and special assumption. However, it is common to assume that there are constant returns to scale in production. Since labour is the only input here, that means that the marginal and average product of labour are both constant. The normalisation of this input/output coefficient to 1 can be achieved by choosing units.
6 Since there are two goods (C, L), there is really only one price, and we can choose either W or P as the numeraire, and set it to 1.
7 This is the solution to maximising $C^{\alpha}L^{(1-\alpha)}$ subject to $C = w\cdot(1-L)+\Pi-T$.
8 Although the original papers were written with overlapping wage contracts, the analysis applies with price-setting firms as well.
9 The astute reader will note that I am assuming that the hill is *smooth*, i.e.that it does not have a corner at the top, as in the case (for example) of a pyramid.
10 I leave it to the reader's imagination to wonder what the envelopes have to do with mountains. However, you can also look it up. Sometimes the envelope theorem is also called the Theorem of the Maximum.
11 I have drawn it here as a big shift simply to make Figure 4.8 clearer, so it should not be taken as drawn 'to scale' with Figure 4.7.
12 I do not really agree with this interpretation: see Chapter 7.
13 Like many solutions in dynamic optimisation, this is not a general result in a mathematical sense, but economists (and engineers) usually assume that the world is sufficiently as it needs to be for this rule to be optimal.
14 This is called a 'one-sided' (S, s) rule, because only one barrier is ever hit: inflation means that the optimal price P^* is always rising, and the problem that the firm faces

is that once it has set its price, its real value is falling due to background inflation until it changes its price again.

15 I say 'seem', because there is a distinction between list prices (the advertised prices) and transactions prices (the prices at which the goods are actually sold). Obviously, discounts given to customers are hard to observe by an outside observer, but they clearly happen (in some markets a discount is expected – for example in the UK car market, no one expects to pay the list price).

16 The Roman setting is needed because an example with 12 months is slightly more complex.

17 The astute reader (or the aspirant bank clerk) will have noticed that I am ignoring the compounding of interest rates over the 10 months. The real annual inflation rate would be a little over 11 per cent if the monthly was 1 per cent. However, in the interested of keeping things simple, I return to the main text.

18 Note that this equates λ to marginal utility.

19 These rays are of course the Income Expansion paths for consumption and leisure.

20 The reader may need to be reminded at this point that although we talk about 'the' household, and 'the' firm, output and so on, this is just a simplifying device: the model is valid with many households, many markets and many goods.

21 This term is sometimes called the Lerner index of monopoly, after the economist Abba Lerner who invented it in 1933.

22 Profts are $\Pi = (P - W)\cdot N = (1 - W)\cdot N = \mu\cdot N$

23 We can find this by differentiating (4.22) with respect to μ, in which case we obtain:

$$\frac{dN}{d\mu} = \frac{-\alpha(1-\alpha)(1-g)}{(1-\alpha\mu)^2}$$

which is negative since for N to be positive, $g < 1$.

24 The multiplier is the sum of the infinite geometric series $dg(1-\alpha)[1 + \alpha\mu + (\alpha\mu)^2 + (\alpha\mu)^3 + \ldots]$.

25 On the details of the Cambridge soap, see Marris (1991, pp. 181–7).

26 Although, of course there were several people who recognised the importance of imperfect competition and macroeconomics; see Dixon and Rankin (1995, pp. 3–5).

27 In my own case, it is really the latter: my original (1987c) paper stressed the Walrasian rather than the Keynesian features of the multpliers.