Chapter 19

The theory of effective demand

> In essence, the "Keynesian revolution" was a shift of emphasis from one type of short-run equilibrium to another type as providing the appropriate theory for actual unemployment situations.
>
> −Edmund Malinvaud (1977), p. 29.

In this and the following chapters the focus is shifted from long-run macroeconomics to short-run macroeconomics. The long-run models concentrated on factors of importance for the economic evolution over a time horizon of at least 10-15 years. With such a horizon the supply side (think of capital accumulation, population growth, and technological progress) is the primary determinant of cumulative changes in output and consumption − the trend. Mainstream macroeconomists see the demand side and monetary factors as of key importance for the *fluctuations* of output and employment about the trend. In a long-run perspective these fluctuations are of only secondary quantitative importance. The preceding chapters have chiefly ignored them. But within shorter horizons, fluctuations are the focal point and this brings the demand-side, monetary factors, market imperfections, nominal rigidities, and expectation errors to the fore. The present and subsequent chapters deal with the role of these short- and medium-run factors for the failure of the laissez-faire market economy to ensure full employment and for the possibilities of active macro policies as a means to improve outcomes.

This chapter introduces building blocks of Keynesian theory of the short run. By "Keynesian theory" we mean a macroeconomic framework that (a) aims at understanding "what determines the actual employment of the available resources",[1] including understanding why mass unemployment arises from time to time, and (b) in this endeavor ascribes a primary role to aggregate demand. Whether a particular building block in this framework comes from Keynes himself, post-war Keynesians, or "new" Keynesians of some sort is not our concern.

[1] Keynes, 1936, p. 4.

We present the basic concept of *effective demand* and compare with the pre-Keynesian (Walrasian) macroeconomic theory which did not distinguish systematically between ex ante demand and supply on the one hand and actual transactions on the other. Attention to this distinction leads to a refutation of *Say's law*, the doctrine that "supply creates its own demand". Next we present some microfoundation for the notion of *nominal price stickiness.* In particular the *menu cost theory* is discussed. We also address the conception of "abundant capacity" as the prevailing state of affairs in an industrialized market economy.

19.1 Stylized facts about the short run

The idea that prices of most goods and services are sticky in the short run rests on the empirical observation that in the short run firms in the manufacturing and service industries typically let output do the adjustment to changes in demand while keeping prices unchanged. In industrialized societies firms are able to do that because under "normal circumstances" there is "abundant production capacity" available in the economy. Three of the most salient short-run features that arise from macroeconomic time series analysis of industrialized market economies are the following (cf. Blanchard and Fischer, 1989, Christiano et al., 1999):

1) Shifts in aggregate demand (induced by sudden changes in the state of confidence, exports, fiscal or monetary policy, or other events) are largely accommodated by changes in quantities rather than changes in nominal prices − *nominal price insensitivity.*

2) Even large movements in quantities are often associated with little or no movement in relative prices − *real price insensitivity.* The real wage, for instance, exhibits such insensitivity in the short run.

3) Nominal prices are *sensitive* to general changes in *input costs.*

These stylized facts pertain to final goods and services. It is not the case that *all* nominal prices in the economy are in the short run insensitive vis-a-vis demand changes. One must distinguish between production of most final goods and services on the one hand and production of primary foodstuff and raw materials on the other. This leads to the associated distinction between "cost-determined" and "demand- determined" prices.

Final goods and services are typically differentiated goods (imperfect substitutes). Their production takes place under conditions of imperfect competition. As a result of existing reserves of production capacity, generally speaking, the *production is elastic w.r.t. demand.* A shift in demand tends to be met by a change in production rather than price. The price changes which do occur are mostly a response to general changes in costs of production. Hence the name "cost-determined" prices.

For *primary foodstuff* and many *raw materials* the situation is different. To increase the supply of most agricultural products requires considerable time. This is also true (though not to the same extent) with respect to mining of raw materials as well as extraction and transport of crude oil. When production is *inelastic* w.r.t. demand in the short run, an increase in demand results in a diminution of stocks and a rise in price. Hence the name "demand-determined prices". The price rise may be enhanced by a speculative element: temporary hoarding in the expectation of further price increases. The price of oil and coffee − two of the most traded commodities in the world market − fluctuate a lot. Through the channel of *costs* the changes in these demand-determined prices spill over to the prices of final goods. Housing construction is time consuming and is also an area where, apart from regulation, demand-determined prices is the rule in the short run.

In industrialized economies manufacturing and services are the main sectors, and the general price level is typically regarded as cost-determined rather than demand determined. Two further aspects are important. First, many wages and prices are set in nominal terms by *price setting agents* like craft unions and firms operating in imperfectly competitive output markets. Second, these wages and prices are in general deliberately kept unchanged for some time even if changes in the environment of the agent occurs; this aspect, possibly due to pecuniary or non-pecuniary costs of changing prices, is known as *nominal price stickiness.* Both aspects have vast consequences for the functioning of the economy as a whole compared with a regime of perfect competition and flexible prices.

Note that *price insensitivity* just refers to the sheer observation of absence of price change in spite of changes in the "environment" − as in the context of facts 1) and 2) above. *Price stickiness* refers to more, namely that prices do not move quickly enough to clear the market in the short run. While price stickiness is in principle a matter of degree, the term includes the limiting case where prices are entirely "fixed" over the period considered − the case of *price rigidity.*

19.2 A simple short-run model

The simple model presented below is close to what Paul Krugman named the *World's Smallest Macroeconomic Model.*[2] The model is crude but nevertheless useful in at least three ways:

[2]Krugman (1999). Krugman tells he learned the model back in 1975 from Robert Hall. As presented here there is an inspiration from Barro and Grossman (1971).

- the model demonstrates the fundamental difference in the *functioning* of an economy with fully flexible prices and one with sticky prices;
- by addressing spillovers across markets, the model is a suitable point of departure for a definition of the Keynesian concept of *effective demand;*
- the model displays the logic behind the Keynesian *refutation* of *Say's law.*

19.2.1 Elements of the model

We consider a monetary closed economy which produces a consumption good. There are three sectors in the economy, a production sector, a household sector, and a public sector with a consolidated government/central bank. Time is discrete. There is a *current period*, of length a month or a quarter of a year, say, and "the future", compressing the next period and onward. Labor is the only input in production. To simplify notation, the model presents its story as if there is just one representative household and one representative firm owned by the household, but the reader should of course imagine that there are numerous agents, all alike, of each kind.

The production function has CRS,

$$Y = AN, \qquad A > 0, \tag{19.1}$$

where Y is aggregate output of a consumption good which is perishable and therefore cannot be stored, A is a technology parameter and N is aggregate employment in the current period. In short- and medium-run macroeconomics the tradition is to use N to denote labor input ("number of hours"), while L is typically used for liquidity demand, i.e., money demand. We follow this tradition.

The price of the consumption good in terms of money, i.e., the *nominal* price, is P. The wage rate in terms of money, the *nominal* wage, is W. We assume that the representative firm maximizes profit, taking these current prices as given. The nominal profit, possibly nil, is

$$\Pi = PY - WN. \tag{19.2}$$

There is free exit from the production sector in the sense that the representative firm can decide to produce nothing. Hence, an equilibrium with positive production requires that profits are non-negative.

The representative household supplies labor inelastically in the amount $\bar{N}$ and receives the profit obtained by the firm, if any. The household demands the consumption good in the amount C^d in the current period (since we want to allow cases of non-market clearing, we distinguish between consumption *demand,*

C^d, and realized consumption, C. Current income not consumed is saved for the future. The output good cannot be stored and there is no loan market. Or we might say that there will exist no interest rate at which intertemporal exchange could be active (neither in the form of a loan market or a market for ownership rights to the firms' profits, if any, in the future). This is because all households are alike, and firms have no use for funding. The only asset on hand for saving is *fiat money* in the form of currency in circulation. Until further notice the money stock is constant.

The preferences of the household are given by the utility function,

$$U = \ln C^d + \beta \ln \frac{\hat{M}}{P^e}, \qquad 0 < \beta < 1, \tag{19.3}$$

where $\hat{M}$ is the amount of money transferred to "the future", and P^e is the expected future price level. The utility discount factor β (equal to $(1+\rho)^{-1}$ if ρ is the utility discount rate) reflects "patience".

Consider the household's choice problem. Facing P and W and expecting that the future price level will be P^e, the household chooses C^d and $\hat{M}$ to maximize U s.t.

$$PC^d + \hat{M} = M + WN + \Pi \equiv B, \qquad N \leq N^s = \bar{N}. \tag{19.4}$$

Here, $M > 0$ is the stock of money held at the beginning of the current period and is predetermined. The actual employment is denoted N and equals the minimum of the amount of employment offered by the representative firm and the labor supply $\bar{N}$ (the principle of voluntary trade). The sum of initial financial wealth, M, and nominal income, $WN + \Pi$, constitutes the budget, B.[3] Payments occur at the end of the period. Nominal financial wealth at the beginning of the next period is $\hat{M} = M + WN + \Pi - PC^d$, i.e., the sum of initial financial wealth and planned saving where the latter equals $WN + \Pi - PC^d$. The benefit obtained by transferring $\hat{M}$ depends on the expected purchasing power of $\hat{M}$, hence it is $\hat{M}/P^e$ that enters the utility function. (Presumably, the household has expectations about real labor and profit income also in the future. But these future incomes are assumed given. So there is no role for changed expectations about the future.)

Substituting $\hat{M} = B - PC^d$ into (19.3), we get the first-order condition

$$\frac{dU}{dC^d} = \frac{1}{C^d} + \beta \frac{P^e}{B - PC^d}(-\frac{P}{P^e}) = 0,$$

[3] As time is discrete, expressions like $M + WN + \Pi$ are legitimate. Although it is meaningless to add a stock and a flow (since they have different denominations), the sum $M + WN + \Pi$ should be interpreted as $M + (WN + \Pi)\Delta t$, where Δt is the period length. With the latter being the time unit, we have $\Delta t = 1$.

which gives

$$PC^d = \frac{1}{1+\beta}B. \tag{19.5}$$

We see that the marginal (= average) propensity to consume is $(1+\beta)^{-1}$, hence inversely related to the patience parameter β. The planned stock of money to be held at the end of the period is

$$\hat{M} = (1 - \frac{1}{1+\beta})B = \frac{\beta}{1+\beta}B.$$

So, the expected price level, P^e, in the future does not affect the demands, C^d and $\hat{M}$. This is a special feature caused by the additive-logarithmic specification of the utility function in (19.3). Indeed, with this specification the substitution and income effects on current consumption of a change in the expected real gross rate of return, $(1/P^e)/(1/P)$, on saving exactly offset each other. And there is no wealth effect on current consumption from a change in the expected rate of return because there is no channel for (or interest in) intertemporal transfer of purchasing power.

Inserting (19.4) and (19.2) into (19.5) gives

$$C^d = \frac{B}{P(1+\beta)} = \frac{M + WN + \Pi}{P(1+\beta)} = \frac{\frac{M}{P} + Y}{1+\beta}, \tag{19.6}$$

In our simple model output demand is the same as the consumption demand C^d. So *clearing* in the output market, in the sense of equality between demand and actual output, requires $C^d = Y$. So, *if* this clearing condition holds, substituting into (19.6) gives the relationship

$$Y = \frac{M}{\beta P}. \tag{19.7}$$

This is only a *relationship* between Y and P, not a solution for any of them since both are endogenous variables so far. Moreover, the relationship is *conditional on clearing* in the output market.

We have assumed that agents take prices as given when making their demand and supply decisions. But we have said nothing about whether nominal prices are flexible or rigid as seen from the perspective of the system as a whole.

19.2.2 The case of competitive markets with fully flexible W and P

What Keynes called "classical economics" is nowadays also often called "Walrasian macroeconomics" (sometime just "pre-Keynesian macroeconomics"). In this theoretical tradition both wages and prices are assumed fully flexible and all markets perfectly competitive.

Firms' ex ante output supply conditional on a hypothetical wage-price pair (W, P) and the corresponding labor demand will be denoted Y^s and N^d, respectively. As we know from microeconomics, the pair (Y^s, N^d) need not be unique, it can easily be a "set-valued function" of (W, P). Moreover, with constant returns to scale in the production function, the range of this function may for certain pairs (W, P) include (∞, ∞).

The distinguishing feature of the Walrasian approach is that wages and prices are assumed fully flexible. Both W and P are thought to adjust immediately so as to clear the labor market and the output market like in a centralized auction market. Clearing in the labor market requires that W and P are adjusted so that actual employment, N, equals labor supply, N^s, which is here inelastic at the given level $\bar{N}$. So

$$N = N^s = \bar{N} = N^d, \tag{19.8}$$

where the last equality indicates that this employment level is willingly demanded by the firms.

We have assumed a constant-returns-to-scale production function (19.1). Hence, the clearing condition (19.8) requires that firms have zero profit. In turn, by (19.1) and (19.2), zero profit requires that the real wage equals labor productivity:

$$\frac{W}{P} = A. \tag{19.9}$$

With clearing in the labor market, output must equal full-employment output,

$$Y = A\bar{N} \equiv Y^f = Y^s, \tag{19.10}$$

where the superscript "f" stands for "full employment", and where the last equality indicates that this level of output is willingly supplied by the firms. For this level of output to match the demand, C^d, coming from the households, the price level must be

$$P = \frac{M}{\beta Y^f} \equiv P^c, \tag{19.11}$$

in view of (19.7) with $Y = Y^f$. This price level is the *classical equilibrium price*, hence the superscript "c". Substituting into (19.9) gives the *classical equilibrium wage*

$$W = AP^c \equiv W^c. \tag{19.12}$$

For general equilibrium we also need that the desired money holding at the end of the period equals the available money stock. By *Walras' law* this equality follows automatically from the household's *Walrasian* budget constraint and clearing in the output and labor markets. To see this, note that the *Walrasian* budget constraint is a *special case* of the budget constraint (19.4), namely the case

$$PC^d + \hat{M} = M + WN^s + \Pi^c, \tag{19.13}$$

where Π^c is the notional profit associated with the hypothetical production plan (Y^s, N^d), i.e.,

$$\Pi^c \equiv PY^s - WN^d. \tag{19.14}$$

The Walrasian budget constraint thus *imposes* replacement of the term for *actual* employment, N, with the households' desired labor supply, $N^s (= \bar{N})$. It also *imposes* replacement of the term for *actual* profit, Π, with the hypothetical profit Π^c ("c" for "classical") calculated on the basis of the firms' aggregate production plan (Y^s, N^d).

Now, let the Walrasian auctioneer announce an arbitrary price vector $(W, P, 1)$, with $W > 0$, $P > 0$, and 1 being the price of the numeraire, money. Then the values of excess demands add up to

$$\begin{aligned} & W(N^d - N^s) + P(C^d - Y^s) + \hat{M} - M \\ = \; & WN^d - PY^s + PC^d + \hat{M} - M - WN^s \quad \text{(by rearranging)} \\ = \; & WN^d - PY^s + \Pi^c \quad \text{(by (19.13))} \\ = \; & WN^d - PY^s + \Pi^c \equiv 0. \qquad \text{(from definition of } \Pi^c \text{ in (19.14))} \end{aligned}$$

This exemplifies *Walras' law,* saying that with Walrasian budget constraints the aggregate value of excess demands is *identically* zero. Walras' law reflects that when households satisfy their Walrasian budget constraint, then as an arithmetic necessity the economy as a whole has to satisfy an aggregate budget constraint for the period in question. It follows that the equilibrium condition $\hat{M} = M$ is ensured as soon as there is clearing in the output and labor markets. And more generally: if there are n markets and $n-1$ of these clear, so does the n'th market.

Consequently, when $(W, P) = (W^c, P^c)$, all markets clear in this flexwage-flexprice economy with perfect competition and a representative household with the "endowment"-pair $(M, \bar{N})$. Such a state of affairs is known as a *classical* or *Walrasian equilibrium.*[4] A key feature is expressed by (19.8) and (19.10): output and employment are *supply-determined*, i.e., determined by the supply of production factors, here labor.

The intuitive mechanism behind this equilibrium is the following adjustment process. Imagine that in an ultra-short sub-period $W/P - A \neq 0$. In case $W/P - A > 0$ (< 0), there will be excess supply (demand) in the labor market. This drives

[4] To underline its one-period nature, it may be called a Walrasian *short-run* or a Walrasian *temporary equilibrium.*

W down (up). Only when $W/P = A$ and full employment obtains, can the system be at rest. Next imagine that $P - P^c \neq 0$. In case $P - P^c > 0$ (< 0), there is excess supply (demand) in the output market. This drives P down (up). Again, only when $P = P^c$ *and* $W/P = A$ (whereby $W = W^c$), so that the output market clears under full employment, will the system be at rest.

This adjustment process is fictional, however. Outside equilibrium the Walrasian supplies and demands, which supposedly drive the adjustment, are artificial constructs. Being functions only of initial resources and price signals, the Walrasian supplies and demands are mutually inconsistent outside equilibrium and can therefore not tell what quantities will be traded during an adjustment process. The story needs a considerable refinement unless one is willing to let the mythical "Walrasian auctioneer" enter the scene and bring about adjustment toward the equilibrium prices without allowing trade until these prices are found.

Anyway, assuming that Walrasian equilibrium has been attained, by *comparative statics* based on (19.11) and (19.12) we see that in the classical regime: (a) P and W are proportional to M; (b) output is at the unchanged full-employment level whatever the level of M. This is the *neutrality of money* result of classical macroeconomics.

The neutrality result also holds when we consider an actual change in the money stock at the beginning of the period. Suppose the government/central bank decides a lump-sum transfer to the households in the total amount $\Delta M > 0$ at the beginning of the period. There is no taxation, and so this implies a government budget deficit which is thus fully financed by money issue.[5] So (19.4) is replaced by

$$PC^d + \hat{M} = M + \Delta M + W\bar{N} + \Pi^c. \tag{19.15}$$

If we replace M in the previous formulas by $M' \equiv M + \Delta M$, we see that money neutrality still holds. As *saving* is income minus consumption, there is now positive nominal private saving of size $S^p = \Delta M + W\bar{N} + \Pi^c - PC^d = M' - M = \Delta M$. On the other hand the government dissaves, in that its saving is $S^g = -\Delta M$, where ΔM is the government budget deficit. So national saving is and remains $S \equiv S^p + S^g = 0$ (it *must* be nil since there are no durable produced goods).

[5] Within the model this is in fact the only way to increase the money stock. As money is the only asset in the economy, a change in the money stock can not be brought about through open-market operations where the central bank buys or sells another financial asset. So $\Delta M > 0$ represents a combination of fiscal policy (the transfers) and monetary policy (the financing of the transfers by money).

19.2.3 The case of imperfect competition and W and P fixed in the short run

In standard Keynesian macroeconomics nominal wages are considered predetermined in the short run, fixed in advance by wage bargaining between workers and employers (or workers' unions and employers' unions). Those who end up unemployed in the period do not try to − or are not able to − undercut those employed, at least not in the current period.

Likewise, nominal prices are set in advance by firms facing downward-sloping demand curves. It is understood that there is a large spectrum of differentiated products, and Y and C are composites of these. This heterogeneity ought of course be visible in the model − and it will become so in Section 19.3. But at this point the model takes an easy way out and ignores the involved aggregation issue.

Let W in the current period be given at the level $\bar{W}$. Because firms have market power, the profit-maximizing price involves a mark-up on marginal cost, $\bar{W}N/Y = \bar{W}/A$ (which is also the average cost). We assume that the price setting occurs under circumstances where the chosen mark-up becomes a *constant* $\mu > 0$, so that

$$P = (1+\mu)\frac{\bar{W}}{A} \equiv \bar{P}. \tag{19.16}$$

While $\bar{W}$ is considered exogenous (not determined within the model), $\bar{P}$ is endogenously determined by the given $\bar{W}$, A, and μ. There are barriers to entry in the short run.

Because of the fixed wage and price, the distinction between *ex ante* (also called *planned* or *intended*) demands and supplies and the *ex post* carried out purchases and sales are now even more important than before. This is because the different markets may now also *ex post* feature excess demand or excess supply (to be defined more precisely below). According to the principle that no agent can be forced to trade more than desired, the actual amount traded in a market must equal the minimum of demand and supply. So in the output market and the labor market the actual quantities traded will be

$$Y = \min(Y^d, Y^s) \quad \text{and} \tag{19.17}$$

$$N = \min(N^d, N^s), \tag{19.18}$$

respectively, where the superscripts "*d*" and "*s*" are now used for demand and supply in a *new* meaning to be defined below. This principle, that the short side of the market determines the traded quantity, is known as the *short-side rule.* The other side of the market is said to be *quantity rationed* or just *rationed* if there is discrepancy between Y^d and Y^s. In view of the produced good being non-storable, intended inventory investment is ruled out. Hence, the firms try to avoid producing more than can be sold. In (19.17) we have thus identified the traded quantity with the produced quantity, Y.

But what exactly do we mean by "demand" and "supply" in this context where market clearing is not guaranteed? We mean what is appropriately called the *effective demand* and the *effective supply* ("effective" in the meaning of "operative" in the market, though possibly frustrated in view of the short-side rule). To make these concepts clear, we need first to define an agent's *effective* budget constraint:

DEFINITION 1 An agent's (typically a household's) *effective budget constraint* is the budget constraint conditional on the perceived price and quantity signals from the markets.

It is the last part, "and quantity signals from the markets", which is not included in the concept of a Walrasian budget constraint. The perceived quantity signals are in the present context a) the *actual* employment constraint faced by the household and b) the profit expected to be received from the firms and determined by their *actual* production and sales. So the household's effective budget constraint is given by (19.4). In contrast, the Walrasian budget constraint is not conditional on quantity signals from the markets but only on the "endowment" $(M, \bar{N})$ and the perceived price signals and profit.

DEFINITION 2 An agent's *effective demand* in a given market is the amount the agent *bids for* in the market, conditional on the perceived price and quantity signals that constrains the bidding.

By "bids for" is meant that the agent is both *able* to buy the amount in question and *wishes* to buy it, given the effective budget constraint. Summing over all potential buyers, we get the *aggregate effective demand* in the market.

DEFINITION 3 An agent's *effective supply* in a given market is the amount the agent *offers for sale* in the market, conditional on perceived price and quantity signals that constrains the offer.

By "offers for sale" is meant that the agent is both *able* to bring that amount to the market and *wishes* to sell it, given the set of opportunities available. Summing over all potential sellers, we get the *aggregate effective supply* in the market.

When $P = \bar{P}$, the aggregate effective output demand, Y^d, is the same as households' consumption demand given by (19.6) with $P = \bar{P}$, i.e.,

$$Y^d = C^d = \frac{\frac{M}{\bar{P}} + Y}{1 + \beta}. \tag{19.19}$$

In view of the inelastic labor supply, households' aggregate effective labor supply is simply

$$N^s = \bar{N}.$$

Firms' aggregate effective output supply is

$$Y^s = Y^f \equiv A\bar{N}. \tag{19.20}$$

Indeed, in the aggregate the firms are *not able* to bring more to the market than full-employment output , Y^f. And every individual firm is not able to bring to the market than what can be produced by "its share" of the labor force. On the other hand, because of the constant marginal costs, every unit sold at the preset price adds to profit. The firms are therefore happy to satisfy any output demand forthcoming − which is in practice testified by a lot of sales promotion.

Firms' aggregate effective demand for labor is constrained by the perceived output demand, Y^d, because the firm would loose by employing more labor. Thus,

$$N^d = \frac{Y^d}{A}. \tag{19.21}$$

By the short-side rule (19.17), combined with (19.20), follows that actual aggregate output (equal to the quantity traded) is

$$Y = \min(Y^d, Y^f) \leqq Y^f.$$

So the following three mutually exclusive cases exhaust the possibilities regarding aggregate output:

$$\begin{aligned} Y &= Y^d < Y^f \quad \text{(the Keynesian regime)}, \\ Y &= Y^f < Y^d \quad \text{(the repressed inflation regime)}, \\ Y &= Y^d = Y^f \quad \text{(the border case)}. \end{aligned}$$

The Keynesian regime: $Y = Y^d < Y^f$.

In this regime we can substitute $Y = Y^d$ into (19.19) and solve for Y:

$$Y = Y^d = \frac{M}{\beta\bar{P}} \equiv Y^k < Y^f \equiv \frac{M}{\beta P^c} = Y^s. \tag{19.22}$$

where we have denoted the resulting output Y^k (the superscript "k" for "Keynesian"). The inequality in (19.22) is required by the definition of the Keynesian regime, and the identity comes from (19.11). Necessary and sufficient for the inequality is that $\bar{P} > P^c \equiv W^c/A$. In view of (19.16), the economy is thus in the Keynesian regime if and only if

$$\bar{W} > W^c/(1+\mu). \tag{19.23}$$

Since $Y < Y^s$ in this regime, we may say there is "excess supply" in the output market or, with a perhaps better term, there is a "buyers' market" situation (sale less than desired). The reservation regarding the term "excess supply" is due to the fact that we should not forget that $Y - Y^s < 0$ is a completely voluntary state of affairs on the part of the price-setting firms.

From (19.1) and the short-side rule now follows that actual employment will be

$$N = N^d = \frac{Y}{A} = \frac{M}{A\beta\bar{P}} < \bar{N} = N^s. \tag{19.24}$$

Also the labor market is thus characterized by "excess supply" or a "buyers' market" situation. Profits are $\Pi = \bar{P}Y - \bar{W}N = (1 - \bar{W}/(\bar{P}A))\bar{P}Y = (1-(1+\mu)^{-1})\beta^{-1}M > 0$, where we have used, first, $Y = AN$, then the price setting rule (19.16), and finally (19.22).

This solution for (Y, N) is known as a *Keynesian equilibrium* for the current period. It is named an *equilibrium* because the system is "at rest" in the following sense: (a) agents do the best they can given the constraints (which include the preset prices and the quantities offered by the other side of the market); and (b) the chosen actions are *mutually compatible* (purchases and sales match). The term equilibrium is here not used in the Walrasian sense of market clearing through instantaneous price adjustment but in the sense of a *Nash equilibrium* conditional on perceived price and quantity signals. To underline its temporary character, the equilibrium may be called a Keynesian *short-run* (or *temporary*) equilibrium. The flavor of the equilibrium is *Keynesian* in the sense that there is unemployment and at the same time it is aggregate demand in the output market, not the real wage, which is the binding constraint on the employment level. A higher propensity to consume (lower discount factor β) results in higher aggregate demand, Y^d, and thereby a higher equilibrium output, Y^k. In contrast, a lower real wage due to either a higher mark-up, μ, or a lower marginal (= average) labor productivity, A, does *not* result in a higher Y^k. On the contrary, Y^k becomes *lower*, and the causal chain behind this goes via a higher $\bar{P}$, cf. (19.16) and (19.22). In fact, the given real wage, $\bar{W}/\bar{P} = A/(1+\mu)$, is consistent with unemployment as well as full employment, see below. It is the sticky nominal *price* at an excessive level, caused by a sticky nominal *wage* at an "excessive" level, that makes unemployment prevail through a too low aggregate demand, Y^d. A lower nominal wage would imply a lower $\bar{P}$ and thereby, for a given M, stimulate Y^d and thus raise Y^k.

In brief, the Keynesian regime leads to an equilibrium where output as well as employment are *demand-determined.*

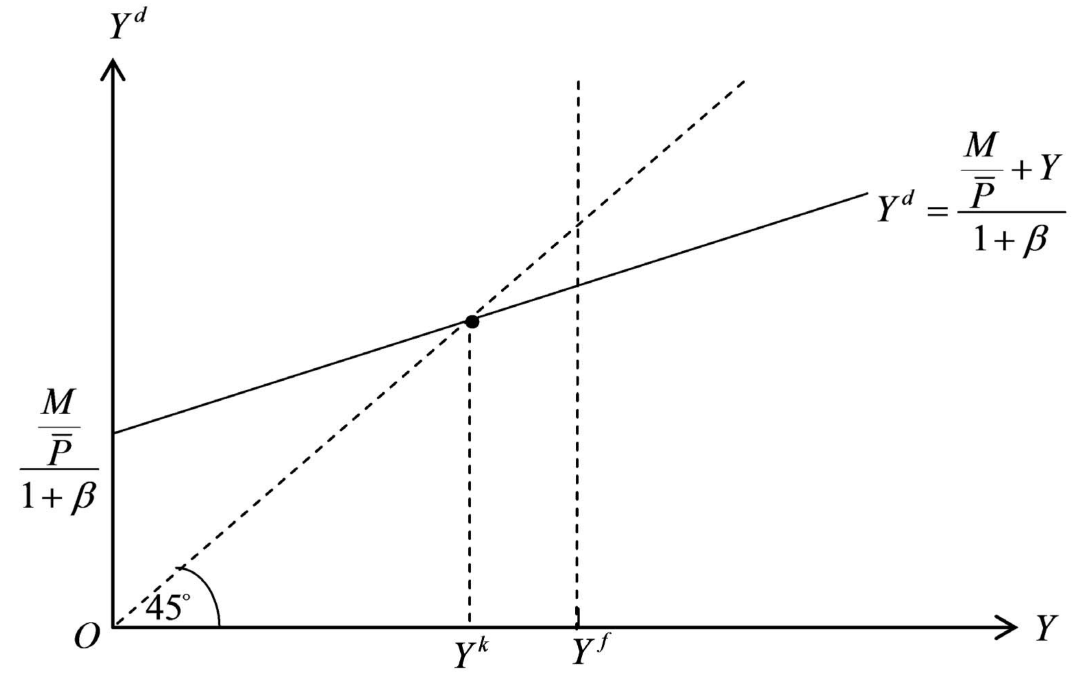

Figure 19.1: The Keynesian regime ($\bar{P} = (1+\mu)\bar{W}/A$, $\bar{W} < W^c/(1+\mu)$, M, and Y^f given).

The "Keynesian cross" and effective demand The situation is illustrated by the "Keynesian cross" in the (Y, Y^d) plane shown in Fig. 19.1, where $Y^d = C^d = (1+\beta)^{-1}(M/\bar{P} + Y)$. We see the vicious circle: Output is below the full-employment level because of low consumption demand; and consumption demand is low because of the low employment. The economy is in a *unemployment trap*. Even though at Y^k we have $\Pi > 0$ and there are constant returns to scale, the individual firm has no incentive to increase production because the firm already produces as much as it rightly perceives it can sell at its preferred price. We also see that here money is *not neutral*. For a given $W = \bar{W}$, and thereby a given $P = \bar{P}$, a higher M results in higher output and higher employment.

Although the microeconomic background we have alluded to is a specific "market power story" (one with differentiated goods and downward sloping demand curves), the Keynesian cross in Fig. 19.1 may turn up also for other microeconomic settings. The key point is the fixed $\bar{P} > P^c$ and fixed $\bar{W} < A\bar{P}$.

The fundamental difference between the Walrasian and the present framework is that the latter allows trade outside Walrasian equilibrium. In that situation the households' consumption demand depends *not* on how much labor the households would *prefer* to sell at the going wage, but on how much they are *able* to sell, that is, on a *quantity signal* received from the labor market. Indeed, it is the *actual* employment, N, that enters the operative budget constraint, (19.4), not the desired employment as in classical or Walrasian theory.

The repressed-inflation regime: $Y = Y^f < Y^d$.

This regime represents the "opposite" case of the Keynesian regime and arises if and only if the opposite of (19.23) holds, namely

$$\bar{W} < W^c/(1+\mu).$$

In view of (19.16), this inequality is equivalent to $\bar{P} < W^c/A \equiv P^c$. Hence $M/(\beta\bar{P}) > M/(\beta P^c) = Y^f = A\bar{N}$. In spite of the high output demand, the shortage of labor hinders the firms to produce more than Y^f. With $Y = Y^f$, output demand, which in this model is always the same as consumption demand, C^d, is, from (19.6),

$$Y^d = \frac{\frac{M}{P} + Y^f}{1+\beta} > Y = Y^s = Y^f. \tag{19.25}$$

As before, effective output supply, Y^s, equals full-employment output, Y^f.

The new element here in that firms perceive a demand level in excess of Y^f. As the real-wage level does not deter profitable production, firms would thus prefer to employ people up to the point where output demand is satisfied. But in view of the short side rule for the labor market, actual employment will be

$$N = N^s = \bar{N} < N^d = \frac{Y^d}{A}.$$

So there is excess demand in both the output market and the labor market. Presumably, these excess demands generate pressure for wage and price increases. By assumption, these potential wage and price increases do not materialize until possibly the next period. So we have a *repressed-inflation equilibrium* (Y, N) $= (Y^f, \bar{N})$, although possibly short-lived.

Fig. 19.2 illustrates the repressed-inflation regime. In the language of the microeconomic theory of quantity rationing, consumers are quantity rationed in the goods market, as realized consumption $= Y = Y^f < Y^d =$ consumption demand. Firms are quantity rationed in the labor market, as $N < N^d$. This is the background for the parlance that in the repressed inflation regime, output and employment are not demand-determined but *supply-determined*. Both the output market and the labor market are *sellers' markets* (purchases less than desired). Presumably, the repressed inflation regime will not last long unless there are wage and price controls imposed by the government, as for instance may be the case for an economy in a war situation.[6]

[6]As another example of repressed inflation (simultaneous excess demand for consumption

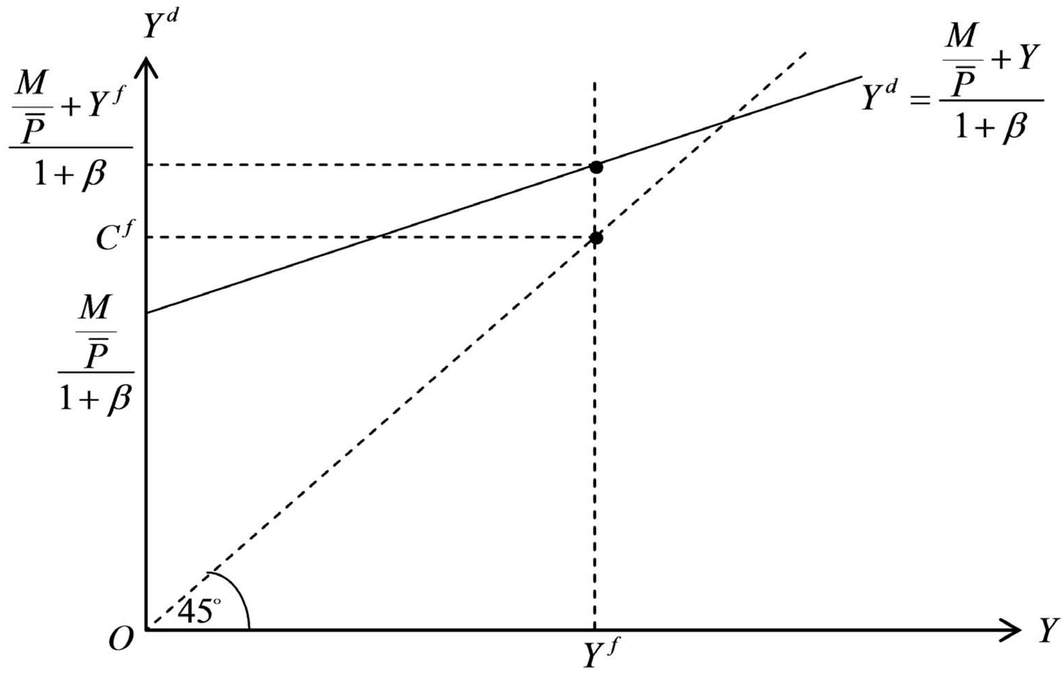

Figure 19.2: The repressed-inflation regime ($\bar{P} = (1 + \mu)\bar{W}/A$, $\bar{W} > W^c/(1 + \mu)$, M, and Y^f given).

The border case between the two regimes: $Y = Y^d = Y^f$.

This case arises if and only if $\bar{W} = W^c/(1+\mu)$, which is in turn equivalent to $\bar{P} = (1+\mu)\bar{W}/A = W^c/A \equiv P^c \equiv M/(\beta Y^f)$. No market has quantity rationing and we may speak of both the output market and the labor market as *balanced markets*.

There are two differences compared with the classical equilibrium, however. The first is that due to market power, there is a wedge between the real wage and the marginal productivity of labor. In the present context, though, where labor supply is inelastic, this does not imply inefficiency but only a higher profit/wage-income ratio than under perfect competition (where the profit/wage-income ratio is zero). The second difference compared with the classical equilibrium is that due to price stickiness, the impact of shifts in exogenous variables will be different. For instance a lower M will here result in unemployment, while in the classical model it will just lower P and W and not affect employment.

In terms of effective demands and supplies Walras' law does not hold

As we saw above, with *Walrasian* budget constraints, the aggregate value of excess demands in the given period is zero for any given price vector, $(W, P, 1)$, with $W > 0$ and $P > 0$. In contrast, with *effective* budget constraints, effective demands and supplies, and the short-side rule, this is no longer so. To see this, consider a pair (W, P) where $W < PA$ and $P \neq P^c \equiv M/(\beta Y^f)$. Such a pair leads to either the Keynesian regime or the repressed-inflation regime. The pair *may*, but need not, equal one of the pairs $(\bar{W}, \bar{P})$ considered above in Fig. 19.1 or 19.2 (we say "need not", because the particular μ-markup relationship between W and P is not needed). We have, *first*, that in both the Keynesian and the repressed-inflation regime, effective output supply equals full-employment output,

$$Y^s = Y^f. \tag{19.26}$$

The intuition is that in view of $W < PA$, the representative firm *wishes* to satisfy any output demand forthcoming but it is only able to do so up to the point of where the availability of workers becomes a binding constraint.

Second, the aggregate value of excess effective demands is, for the considered

goods and labor) we may refer to Eastern Europe before the dissolution of the Soviet Union in 1991. In response to severe and long-lasting rationing in the consumption goods markets, households tended to decrease their labor supply (Kornai, 1979). This example illustrates that if labor supply is elastic, the *effective* labor supply may be less than the Walrasian labor supply due to spillovers from the output market.

price vector $(W, P, 1)$, equal to

$$\begin{aligned}
& W(N^d - N^s) + P(C^d - Y^s) + \hat{M} - M \\
= \; & W(N^d - \bar{N}) + PC^d + \hat{M} - M - PY^f \\
= \; & W(N^d - \bar{N}) + WN + \Pi - PY^f \quad \text{(by (19.4))} \\
= \; & W(N^d - \bar{N}) + PY - PY^f \quad \text{(by (19.2))} \\
= \; & W(N^d - \bar{N}) + P(Y - Y^f) \begin{cases} < 0 \text{ if } P > M/(\beta Y^f), \text{ and} \\ > 0 \text{ if } P < M/(\beta Y^f) \text{ and } W < PA. \end{cases}
\end{aligned} \tag{19.27}$$

The aggregate value of excess effective demands is thus not identically zero. As expected, it is negative in a Keynesian equilibrium and positive in a repressed-inflation equilibrium.[7] The reason that Walras' law does not apply to effective demands and supplies is that outside Walrasian equilibrium some of these demands and supplies are not realized in the final transactions.

This takes us to Keynes' refutation of Say's law and thereby what Keynes and others regarded as the core of his theory.

Say's law and its refutation

The classical principle "supply creates its own demand" (or "income is automatically spent on products") is named Say's law after the French economist and business man Jean-Baptiste Say (1767-1832). In line with other classical economists like David Ricardo and John Stuart Mill, Say maintained that although mismatch between demand and production can occur, it can only occur in the form of excess production in some industries at the same time as there is excess demand in other industries.[8] General overproduction is impossible. Or, by a classical catchphrase:

> Every offer to sell a good implies a demand for some other good.

By "good" is here meant a produced good rather than just any traded article, including for instance money. Otherwise Say's law would be a platitude (a simple implication of the definition of trade). So, interpreting "good" to mean a produced good, let us evaluate Say's law from the point of view of the result (19.27). We first subtract $W(N^d - N^s) = W(N^d - \bar{N})$ on both sides of (19.27), then insert (19.26) and rearrange to get

$$P(C^d - Y) + \hat{M} - M = 0, \tag{19.28}$$

[7] At the same time, (19.27) together with the general equations $N^d = \bar{N}$ and $Y^s = Y^f$, shows that we have $\hat{M} = M$ in a Keynesian equilibrium (where $Y = C^d$) and $\hat{M} < M$ in a repressed-inflation equilibrium (where $Y = Y^f$).

[8] There were two dissidents at this point, Thomas Malthus (1766–1834) and Karl Marx (1818–1883), two writers that were otherwise not much aggreeing.

for any $P > 0$. Consider the case $W < AP$. In this situation every unit produced and sold is profitable. So any Y in the interval $0 < Y \leq Y^f$ is profitable from the supply side angle. Assume further that $P = \bar{P} > P^c \equiv M/(\beta Y^f)$. This is the case shown in Fig. 19.1. The figure illustrates that aggregate demand *is* rising with aggregate production. So far so well for Say's law. We also see that if aggregate production is in the interval $0 < Y < Y^k$, then C^d $(= Y^d) > Y$. This amounts to excess demand for goods and in effect, by (19.28), excess supply of money. Still, Say's law is not contradicted. But if instead aggregate production is in the interval $Y^k < Y \leq Y^f$, then C^d $(= Y^d) < Y$; now there is *general overproduction.* Supply no longer creates its own demand. There is a general shortfall of demand. By (19.28), the other side of the coin is that when $C^d < Y$, then $\hat{M} > M$, which means excess demand for money. People try to hoard money rather than spend on goods. Both the Great Depression in the 1930s and the Great Recession 2008- can be seen in this light.[9]

The refutation of Say's law does not depend on the market power and constant markup aspects we have adhered to above. All that is needed for the argument is that the agents are price takers within the period. Moreover, the refutation does not hinge on *money* being the asset available for transferring purchasing power from one period to the next. We may imagine an economy where M represents *land* available in limited supply. As land is also a non-produced store of value, the above analysis goes through − with one exception, though. This is that ΔM in (19.15) can no longer be interpreted as a policy choice. Instead, a positive ΔM could be due to discovery of new land.

We conclude that general overproduction is possible, and Say's law is thereby refuted. It might be objected that our "aggregate reply" to Say's law is not to the point since Say had a disaggregate structure with many industries in mind. Considering explicitly a multiplicity of production sectors makes no essential difference, however, as the following example will show.

Many industries* Suppose there is still one labor market, but m industries with production function $y_i = An_i$, where y_i and n_i are output and employment in industry i, respectively, $i = 1, 2, \ldots, m$. Let the preferences of the representative household be given by

$$U = \sum_i \gamma_i \ln c_i + \beta \ln \frac{\hat{M}}{P^e}, \qquad \gamma_i > 0, i = 1, 2, \ldots, m, \quad 0 < \beta < 1.$$

[9] Paul Krugman stated it this way: "When everyone is trying to accumulate cash at the same time, which is what happened worldwide after the collapse of Lehman Brothers, the result is an end to demand [for output], which produces a severe recession" (Krugman, 2009).

In analogy with (19.4), the budget constraint is

$$\sum_i P_i c_i + \hat{M} = B \equiv M + W \sum_i n_i + \sum_i \Pi_i = M + \sum_i P_i y_i,$$

where the last equality comes from

$$\Pi_i = P_i y_i - W n_i.$$

Utility maximization gives $P_i c_i = \gamma_i B/(1+\beta)$.

As a special case, consider $\gamma_i = 1/m$ and $P_i = P$, $i = 1, 2, \ldots, m$. Then

$$c_i = \frac{B/m}{(1+\beta)P}, \tag{19.29}$$

and

$$B = M + P \sum_i y_i \equiv M + PY.$$

Substituting into (19.29), we thus find demand for consumption good i as

$$c_i = \frac{\frac{M/m}{P} + Y/m}{1+\beta} \equiv y^d, \quad \text{for all } i.$$

Let $P > \min\left[W/A, M/(\beta Y^f)\right]$, where $Y^f \equiv A\bar{N}$. It follows that every unit produced and sold is profitable and that

$$my^d = \frac{\frac{M}{P} + Y}{1+\beta} \leq \frac{\frac{M}{P} + Y^f}{1+\beta} < Y^f,$$

where the weak inequality comes from $Y \leq Y^f$ (always) and the strict inequality from $P > M/(\beta Y^f)$.

Now, suppose good 1 is brought to the market in the amount y_1, where $y^d < y_1 < Y^f/m$. Industry 1 thus experiences a shortfall of demand. Will there in turn necessarily be another industry experiencing excess demand? No. To see this, consider the case $y^d < y_i < Y^f/m$ for all i. All these supplies are profitable from a supply side point of view, and enough labor is available. Indeed, by construction the resource allocation is such that

$$my^d < \sum y_i \equiv Y \leq m\bar{y} < Y^f, \tag{19.30}$$

where $\bar{y} = \max[y_1, \ldots, y_m] < Y^f/m$. This is a situation where people try to save (hoard money) rather than spend all income on produced goods. It is an example of *general overproduction,* thus falsifying Say's law.

In the special case where all $y_i = Y/m$, the situation for each single industry can be illustrated by a diagram as that in Fig. 19.1. Just replace Y^d, Y, Y^k, Y^f, and M in Fig. 19.1 by y^d, Y/m, $Y^k/m \equiv M/(m\beta P)$, Y^f/m, and M/m, respectively.

19.2.4 Short-run adjustment dynamics*

We now return to the aggregate setup. Apart from the border case of balanced markets, we have considered two kinds of "fix-price equilibria", *repressed inflation* and *Keynesian equilibrium*. Many economists consider nominal wages and prices to be less sticky upwards than downwards. So a repressed inflation regime is typically regarded as having little durability (unless there are wage and price controls imposed by a government). It is otherwise with the Keynesian equilibrium. A way of thinking about this is the following.

Suppose that up to the current period full-employment equilibrium has applied: $Y = Y^d = M/(\beta\bar{P}) = Y^f$ and $\bar{P} = (1+\mu)\bar{W}/A = W^c/A \equiv P^c \equiv M/(\beta Y^f)$. Then, for some external reason, at the start of the current period a *rise* in the patience parameter occurs, from β to β', so that the new propensity to save is $\beta'/(1+\beta') > \beta/(1+\beta)$. We may interpret this as "precautionary saving" in response to a sudden fall in the general "state of confidence".

Let our "period" be divided into n sub-periods, indexed $i = 0, 1, 2, \ldots, n-1$, of length $1/n$, where n is "large". At least within the first of these sub-periods, the preset $\bar{W}$ and $\bar{P}$ are maintained and firms produce without having yet realized that aggregate demand will be lower than in the previous period. After a while firms realize that sales do not keep track with production.

There are basically two kinds of reaction to this situation. One is that wages and prices are maintained throughout all the sub-periods, while production is gradually scaled down to the Keynesian equilibrium $Y^k = M/(\beta'\bar{P})$. Another is that wages and prices adjust downward so as to soon reestablish full-employment equilibrium. Let us take each case at a time.

Wage and price stay fixed: Sheer quantity adjustment For simplicity we have assumed that the produced goods are perishable. So unsold goods represent a complete loss. If firms fully understand the functioning of the economy and have model-consistent expectations, they will adjust production per time unit down to the level Y^k as fast as possible. Suppose instead that firms have naive adaptive expectations of the form

$$C^e_{i-1,i} = C_{i-1}, \qquad i = 0, 1, 2, ..., n.$$

This means that the "subjective" expectation, formed in sub-period $i-1$, of demand next sub-period is that it will equal the demand in sub-period $i-1$. Let the time-lag between the decision to produce and the observation of the demand correspond to the length of the sub-periods. It is profitable to satisfy demand, hence actual output in sub-period i will be

$$Y_i = C^e_{i-1,i} = C^d_{i-1} = \frac{M/\bar{P}}{1+\beta'} + \frac{Y_{i-1}}{1+\beta'},$$

in analogy with (19.19). This is a linear first-order difference equation in Y_i, with constant coefficients. The solution is (see Math Tools)

$$Y_i = (Y_0 - Y^{*\prime}) \left(\frac{1}{1+\beta'}\right)^i + Y^{*\prime}, \qquad Y^{*\prime} = \frac{M}{\beta'\bar{P}} = Y^k < Y^f. \tag{19.31}$$

Suppose $\beta' = 0.9$, say. Then actual production, Y_i, converges fast towards the steady-state value Y^k. When $Y = Y^k$, the system is at rest. Fig. 19.x illustrates. Although there is excess supply in the labor market and therefore some downward pressure on wages, the Keynesian presumption is that the workers's side in the labor market generally withstand the pressure.[10]

Fig. 19.x about here (not yet available).

The process (19.31) also applies "in the opposite direction". Suppose, starting from the Keynesian equilibrium $Y = M/(\beta'\bar{P})$, a *reduction* in the patience parameter β' occurs, such that $M/(\beta'\bar{P})$ increases, but still satisfies $M/(\beta'\bar{P}) < Y^f$. Then the initial condition in (19.31) is $Y_0 < Y^{*\prime}$, and the greater propensity to consume leads to an upward quantity adjustment.

Downward wage and price adjustment Several of Keynes' contemporaries, among them A. C. Pigou, maintained that the Keynesian state of affairs with $Y = Y^k < Y^f$ could only be very temporary. Pigou's argument was that a fall in the price level would take place and lead to higher purchasing power of M. The implied stimulation of aggregate demand would bring the economy back to full employment. This hypothetically equilibrating mechanism is known as the "real balance effect" or the "Pigou effect" (after Pigou, 1943).

Does the argument go through? To answer this, we imagine that the time interval between different rounds of wage and price setting is as short as our sub-periods. We imagine the time interval between households' decision making to be equally short. Given the fixed markup μ, an initial fall in the preset $\bar{W}$ is needed to trigger a fall in the preset $\bar{P}$. The new *classical* equilibrium price and wage levels will be

$$P^{c\prime} = \frac{M}{\beta' Y^f} \text{ and } W^{c\prime} = AP^{c\prime}.$$

Both will thus be lower than the original ones − by the same factor as the patience parameter has risen, i.e., the factor β'/β. In line with "classical" thinking, assume that soon after the rise in the propensity to save, the incipient unemployment

[10]Possible explanations of downward wage stickiness are discussed in Chapter 24.

prompts wage setters to reduce $\bar{W}$ and thereby price setters to reduce $\bar{P}$. Let both $\bar{W}$ and $\bar{P}$ after a few rounds be reduced by the factor β'/β. Denoting the resulting wage and price $\bar{W}'$ and $\bar{P}'$, respectively, we then have

$$\bar{W}' = \frac{W^{cl}}{1+\mu}, \qquad \bar{P}' = (1+\mu)\frac{\bar{W}'}{A} = \frac{W^{cl}}{A} \equiv P^{cl} \equiv \frac{M}{\beta' Y^f}.$$

Seemingly, this restores aggregate demand at the full-employment level $Y^d = M/(\beta'\bar{P}') = Y^f$.

While this "classical" adjustment is conceivable in the abstract, Keynesians question its practical relevance for several reasons:

1. Empirically, it seems to be particularly in the downward direction that nominal wages are sticky. And without an initial fall in the nominal wage, the downward wage-price spiral does not get started.

2. If downward wage-price spiral does get started, the implied deflation increases the implicit real interest rate, $(P_t - P_{t+1})/P_{t+1}$. In a more elaborate modeling of consumption and investment, this would tend to dampen aggregate demand rather than the opposite.

3. Additional points, when going a little outside the present simple model, are:

    (a) the monetary base is in reality only a small fraction of financial wealth, and so the real balance effect can not be very powerful unless the fall in the price level is drastic;

    (b) many firms and households have nominal debt, the real value of which would rise, thereby potentially leading to bankruptcies and a worsening of the confidence crisis, thus counteracting a return to full employment.

A clarifying remark. In this context we should be aware that there are two kinds of "price flexibility" to be distinguished: "imperfect" versus "perfect" (or "full") price flexibility. The first kind relates to a *gradual* price process, for instance generated by a wage-price spiral as at item 2 above. The latter kind relates to *instantaneous* and complete price adjustment as with a Walrasian auctioneer. It is the first kind that may be destabilizing rather than the opposite.

19.2.5 Digging deeper

As it stands the above theoretical framework has many limitations. The remainder of this chapter gives an introduction to how the following three problems have been dealt with in the literature:

(i) Price setting should be explicitly modeled, and in this connection there should be an explanation of price stickiness.

(ii) It should be made clear how to come from the existence of many differentiated goods and markets with imperfect competition to aggregate output and income which in turn constitute the environment conditioning the individual agents' actions.

(iii) The analysis has ignored that capital equipment is in practice an additional factor constraining production.

In subsequent chapters we consider additional problems:

(iv) Also wage setting should be explicitly modeled, and in this connection there should be an explanation of wage stickiness.

(v) At least one additional financial asset, an interest-bearing asset, should enter. This will open up for intertemporal trade and for clarifying the primary function of money as a medium of exchange rather than as a store of value.

(vi) The model should include forward-looking decision making and endogenous expectations.

(vii) The model should be truly dynamic with gradual wage and price changes depending on the market conditions and expectations. This should lead to an explanation why wages and prices do not tend to find their market clearing levels relatively fast.

The next section deals with point (i) and (ii), and Section 19.4 with point (iii).

19.3 Price setting and menu costs

The classical theory of perfectly flexible wages and prices and neutrality of money treats wages and prices as if they were prices on assets traded in centralized auction markets. In contrast, the Keynesian conception is that the general price level is a weighted average of millions of individual prices set − and sooner or later reset − in an asynchronous way by price setters in a multitude of markets and localities.

What we need to understand the determination of prices and their sometimes slow response to changed circumstances, is a theory of how agents set prices and decide when to change them and by how much. This brings the objectives and constraints of agents with market power into the picture. So *imperfect competition* becomes a key ingredient of the theory.

19.3.1 Imperfect competition with price setters

Suppose the market structure is one with *monopolistic competition*:

1. There is a given "large" number, m, of firms and equally many (horizontally) differentiated products.

2. Each firm supplies its own differentiated product on which it has a monopoly and which is an imperfect substitute for the other products.

3. A price change by one firm has only a negligible effect on the demand faced by any other firm.

Another way of stating property 3 is to say that firms are "small" so that each good constitutes only a small fraction of the sales in the overall market system. Each firm faces a perceived downward-sloping demand curve and chooses a price which maximizes the firm's expected profit, thus implying a mark-up on marginal costs. There is no perceivable reaction from the firm's (imperfect) competitors. So the monopolistic competition setup abstracts from strategic interaction between the firms and is thereby different from oligopoly.

With respect to assets, so far our framework corresponds to the World's Smallest Macroeconomic Model of Section 19.2 in the sense that there are no commercial banks and no other non-human assets than fiat money.

Price setting firms

In the short run there is a given large number, m, of firms and equally many (horizontally) differentiated products. Firm i has the production function $y_i = An_i^{\alpha}$, where n_i is labor input (raw materials and physical capital ignored).[11] For notational convenience we imagine measurement units are such that $A = 1$. Thereby,

$$n_i = y_i^{1/\alpha}, \quad 0 < \alpha \leq 1, \quad i = 1, 2, \ldots, m, \quad m \text{ "large"}. \tag{19.32}$$

To extend the perspective compared with Section 19.2, the possibility of rising marginal costs ($\alpha < 1$) is now included.

The demand constraint faced by the firm ex ante is perceived by the firm to be

$$y_i = \left(\frac{P_i}{P}\right)^{-\eta} \frac{Y^e}{m} \equiv D\left(\frac{P_i}{P}, \frac{Y^e}{m}\right), \qquad \eta > 1, \tag{19.33}$$

[11] The following can be seen as an application of the more general framework with price-setting firms outlined at the end of Chapter 2.

where P_i is the price set by the firm and fixed for some time, P is the "general price level" (taken as given by firm i because it is "small" enough for its price to have any noticeable effect on P), Y^e/m is the expectation (for simplicity the same for all firms) of the position of the demand curve, and η is the (absolute) price elasticity of demand (assumed greater than one since otherwise there is no finite profit maximizing price).[12] The firms' expectation of the position of the demand curve reflects their expectation, Y^e, of the general level of demand in the economy.

Let firm i choose P_i at the end of the previous period with a view to maximization of expected nominal profit in the current period:

$$\max_{P_i} \Pi_i = P_i y_i - W n_i \quad \text{s.t. } (19.33),$$

where W is the going nominal wage, taken as given by the firm. We may substitute (19.32) and the constraint (19.33) into the profit function to get an unconstrained maximization problem which is then solved for P_i. The more intuitive approach, however, is to apply the rule that the profit maximizing quantity of a monopolist (in the standard case with non-decreasing marginal cost) is the quantity at which marginal revenue equals marginal cost,

$$MR_i = MC_i = \frac{W}{\alpha} y_i^{\frac{1}{\alpha}-1}. \tag{19.34}$$

Total revenue is $TR_i = P_i(y_i)y_i$, where $P_i(y_i)$ is the price at which expected sales is y_i units. So

$$\begin{aligned} MR_i &= \frac{dTR_i}{dy_i} = P_i(y_i) + y_i P_i'(y_i) = P_i(y_i)(1 + \frac{y_i P_i'(y_i)}{P_i(y_i)}) \qquad (19.35) \\ &= P_i(y_i)(1 - \frac{1}{\eta}) = \left(\frac{y_i}{Y^e/m}\right)^{-1/\eta} P \frac{\eta - 1}{\eta}, \end{aligned}$$

where we have inserted $P_i(y_i) = (y_i/(Y^e/m))^{-1/\eta} P$, which follows from (19.33). Inserting this into (19.34), the unique solution for y_i is the profit maximizing quantity, given Y^e and P. We denote this planned individual output level y_i^e.

The associated price is

$$\bar{P}_i = P_i(y_i^e) = \frac{\eta}{\eta - 1} \frac{W}{\alpha} (y_i^e)^{\frac{1}{\alpha}-1} \equiv (1+\mu) \frac{W}{\alpha} (y_i^e)^{\frac{1}{\alpha}-1}, \tag{19.36}$$

[12] Chapter 20 goes deeper and gives an account of the class of consumer preferences that underlie the constancy of this price elasticity. That chapter also presents a precise definition of the "general price level".

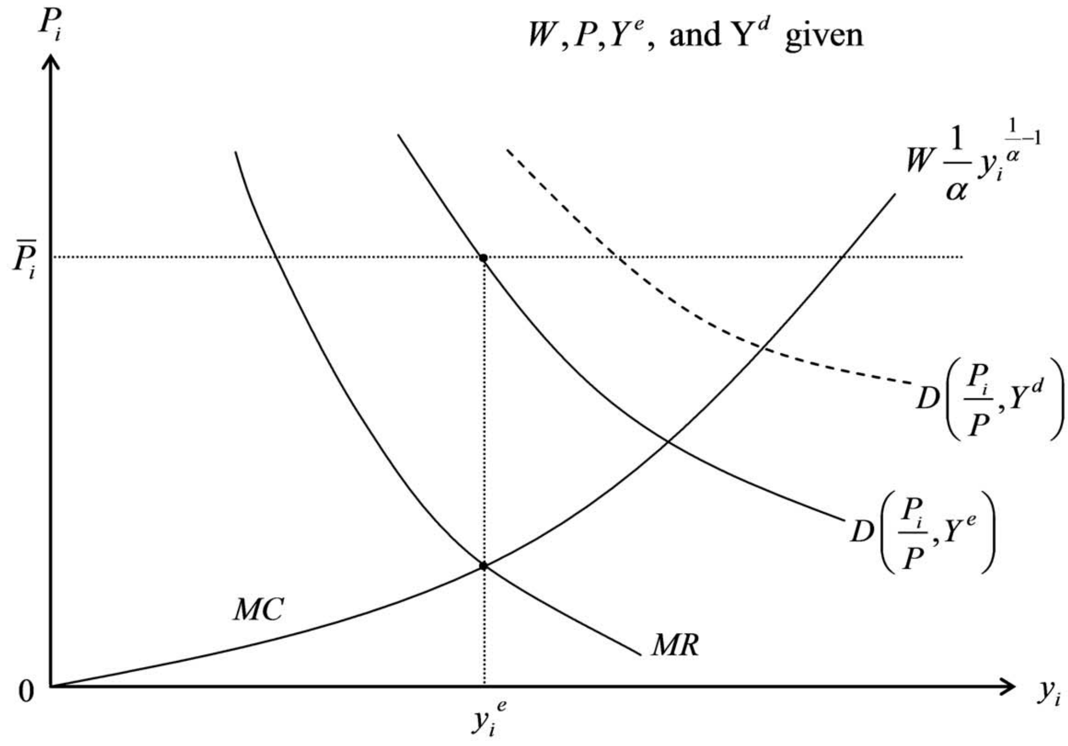

Figure 19.3: Firm i's price choice under the expectation that the general demand level will be Y^ϵ (the case $\alpha < 1$). The demand curve for a higher general demand level, Y^d, is also shown.

where the second equality comes from (19.35) inserted into (19.34), and μ is the mark-up on marginal cost at output level y_i^e, that is, $1 + \mu = \eta(\eta - 1)^{-1}$ $= 1 + (\eta - 1)^{-1}$.

This outcome is illustrated in Fig. 19.3 for the case $\alpha < 1$ (decreasing returns to scale). For fixed Y^e and P, the perceived demand curve faced by firm i is shown as the solid downward-sloping curve $D(P_i/P, Y^e/m)$ to which corresponds the marginal revenue curve, MR. For fixed W, the marginal costs faced by the firm are shown as the upward-sloping marginal cost curve, MC. It is assumed that firm i knows W in advance. The price $\bar{P}_i$ is set in accordance with the rule $MR = MC$.

Because of the symmetric setup, all firms end up choosing the same price, which therefore becomes the general price level, i.e., $\bar{P}_i = P$, $i = 1, 2, \ldots, m$. So all firms' planned level of sales equals the expected average real spending per consumption good, i.e., $y_i^e = Y^e/m \equiv y^e$, $i = 1, 2, \ldots, m$.

In case actual aggregate demand, Y^d, turns out as expected, firm i's actual output, y_i, equals the planned level, y^e. As this holds for all i, we have in this case

$$Y \equiv \frac{\sum_i \bar{P}_i y_i}{P} = \sum_i y_i = \sum_i y^e = Y^e = Y^d. \tag{19.37}$$

In some new-Keynesian models the labor market is described in an analogue way with heterogeneous labor organized in craft unions and monopolistic competition between these. To avoid complicating the exposition, however, we here treat labor as homogeneous. And until further notice we will simply *assume* that at the going wage there is enough labor available to carry out the desired production. We shall consider the question: If aggregate demand in the current period turns out different from expected, what will the firms do: change the price or output or both? To fix ideas we will concentrate on the case where the wage level, W, is unchanged. In that case the answer will be that "only output will be adjusted" if one of the following conditions is present:

(a) The marginal cost curve is horizontal and the price elasticity of demand is constant.

(b) The perceived cost of price adjustment exceeds the potential benefit.

That point (a) is sufficient for "only output will be adjusted" (as long as W is unchanged) follows from (19.36) with $\alpha = 1$. With rising marginal costs ($\alpha < 1$), however, the presence of sufficient price adjustment costs becomes decisive.

19.3.2 Price adjustment costs

The literature has modelled price adjustment costs in two different ways. *Menu costs* refer to the case where there are *fixed costs* of changing price. Another case considered in the literature is the case of *strictly convex adjustment costs*, where the marginal price adjustment cost is increasing in the size of the price change.

As to *menu costs*, the most obvious examples are costs associated with:

1. remarking commodities with new price labels,

2. changing price lists ("menu cards") and catalogues.

But "menu costs" should be interpreted in a broader sense, including pecuniary as well as non-pecuniary costs associated with:

3. information-gathering and recomputing optimal prices,

4. conveying rapidly the new directives to the sales force,

5. the risk of offending customers by frequent price changes (whether these are upward or downward),

6. search for new customers willing to pay a higher price,

7. renegotiation of contracts.

Menu costs induce firms to change prices less often than if no such costs were present. And some of the points mentioned in the list above, in particular point 6 and 7, may be relevant also in labor markets.

The menu cost theory provides the more popular explanation of nominal price stickiness. Another explanation rests on the presumption of *strictly convex price adjustment costs*. In this theory the cost for firm i of changing price is assumed to be $k_{it} = \xi_i(P_{it} - P_{it-1})^2$, $\xi_i > 0$. Under this assumption the firm is induced to avoid *large* price changes, which means that it tends to make frequent, but small price adjustments. This theory is related to the customer market theory. Customers search less frequently than they purchase. A large upward price change may be provocative to customers and lead them to do search in the market, thereby perhaps becoming aware of attractive offers from other stores. The implied "kinked demand curve" can explain that firms are reluctant to suddenly increase their price.[13]

Below we describe the role of the first kind of price adjustment costs, menu costs, in more detail.

The menu cost theory

The menu cost theory originated almost simultaneously in Akerlof and Yellen (1985a, 1985b) and Mankiw (1985). It makes up the predominant microfoundation for the presumption that nominal prices and wages tend to be sticky in the short run vis-a-vis demand changes. For simplicity, we will concentrate on product prices and downplay the intertemporal aspects of price-setting.

The key theoretical insight of the menu cost theory is that even *small* menu costs can be enough to prevent firms from changing their price vis-a-vis demand changes. This is because the opportunity cost of not changing price is only of *second order*, that is, "small", which is a reflection of the *envelope theorem*; hence the potential benefit of changing price can easily be smaller than the cost of changing price. Yet, owing to imperfect competition (price $> MC$), the effect on aggregate output, employment, and welfare of not changing prices is of *first order*, i.e., "large". Let us spell this out in detail.

[13] For details in a macro context, see McDonald (1990).

As in the World's Smallest Macroeconomic Model, suppose the aggregate demand is proportional to the real money stock:

$$Y^d = \frac{M}{\beta P}, \tag{19.38}$$

where $\beta \in (0, 1)$ is a parameter reflecting consumers' patience. Consider now firm i, $i = 1, 2, \ldots, m$, contemplating its pricing policy. With actual aggregate demand as given by (19.38) inserted into (19.33), the nominal profit as a function of the chosen price, P_i, becomes

$$\begin{aligned}\Pi_i &= P_i y_i - W y_i^{1/\alpha} = P_i \left(\frac{P_i}{P}\right)^{-\eta} \frac{M}{m\beta P} - W \left(\left(\frac{P_i}{P}\right)^{-\eta} \frac{M}{m\beta P}\right)^{1/\alpha} \\ &\equiv \Pi(P_i, P, W, M).\end{aligned} \tag{19.39}$$

Suppose that, initially, $P_i = \bar{P}_i$, where $\bar{P}_i$ is the unique price that maximizes Π_i, given P, W, and M. By (19.36) with $y_i^e = M/(m\beta P)$, we have

$$\bar{P}_i = (1 + \mu)\frac{W}{\alpha}\left(\frac{M}{m\beta P}\right)^{\frac{1}{\alpha}-1}. \tag{19.40}$$

In our simplifying setup there is complete symmetry across the firms so that the profit maximizing price is in fact the same for all firms. Nevertheless we maintain the subscript i on the profit-maximizing price since the logic of the menu cost theory is valid independently of this symmetry. We let $\bar{\Pi}_i$ denote firm i's maximized profit, i.e.,

$$\bar{\Pi}_i = \Pi(\bar{P}_i, P, W, M),$$

as illustrated in Fig. 19.4.

In view of the constant price elasticity of demand, η, and hence a constant markup, μ, if marginal costs are constant ($\alpha = 1$), then the profit-maximizing price is unaffected by a change in aggregate demand, cf. (19.40). This is the well-known case where, owing to constancy of marginal costs, The challenging case in a Keynesian context is the case with rising marginal costs. So let us assume that $\alpha < 1$. In this situation, by (19.40), a higher M, for unchanged P and W, will imply higher $\bar{P}_i$ as also illustrated in Fig. 19.4.

Given the price $P_i = \bar{P}_i$, set in advance, suppose that, at the beginning of the period, an unanticipated, fully money-financed lump-sum transfer payment to the households takes place so that M in (19.38) is replaced by $M' = M + \Delta M$, where $\Delta M > 0$. Suppose further that both W and P remain unchanged, that is, no other price setter responds by changing price. Let $\bar{P}_i'$ denote the new price which under these conditions would be profit maximizing for firm i in the absence

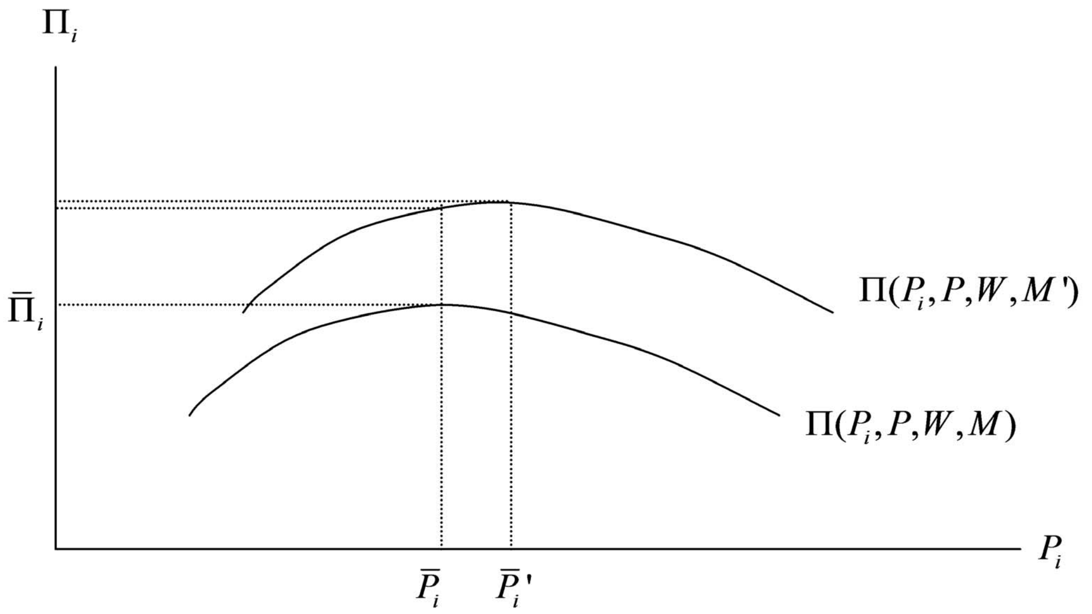

Figure 19.4: The profit curve is flat at the top ($\alpha < 1$, P and W are fixed, $M' > M$).

of menu costs. Fig. 19.4 illustrates. Will firm i have an incentive to change its price to $\bar{P}_i'$? Not necessarily. The menu cost may exceed the opportunity cost associated with not changing price. This opportunity cost to firm i tends to be small. Indeed, considering the marginal effect on Π of the higher M, we have

$$\begin{aligned}\frac{d\Pi}{dM}(\bar{P}_i, P, W, M) &= \frac{\partial \Pi}{\partial P_i}(\bar{P}_i, P, W, M)\frac{\partial P_i}{\partial M} + \frac{\partial \Pi}{\partial M}(\bar{P}_i, P, W, M) \\ &= 0 + \frac{\partial \Pi}{\partial M}(\bar{P}_i, P, W, M).\end{aligned} \tag{19.41}$$

The first term on the right-hand side of (20.36) vanishes at the profit maximum because $\partial \Pi / \partial P_i = 0$ at the point $(\bar{P}_i, P, W, M)$..The profit curve is flat at the profit-maximizing price $\bar{P}_i$. Moreover, since our thought experiment is one where P and W remain unchanged, there is no indirect effect of the rise in M via P or W. Thus, only the direct effect through the fourth argument of the profit function is left. And this effect is independent of a marginal change in the chosen price. This result reflects the *envelope theorem*: in an interior optimum, the total derivative of a maximized function w.r.t. a parameter equals the partial derivative w.r.t. that parameter.[14]

The relevant parameter here is the aggregate money stock, M. As Fig. 19.4 visualizes, the effect of a small change in M on the profit is approximately the

[14]For a general statement, see Math Tools.

same (to a first order) whether or not the firm adjusts its price. In fact, owing to the envelope theorem, for an infinitesimal change in M, the profit of firm i is not affected at all by a marginal change in its price.

For a *finite* change in M this is so only approximately. First, (19.39) shows that the entire profit curve is shifted up, cf. Fig. 19.4. Second, from (19.40) follows that there *will* be a discernible rise in the profit-maximizing price, in Fig. 19.4 from $\bar{P}_i$ to $\bar{P}_i'$. So the new top of the profit curve is north-east of the old. It follows that by not changing price a potential profit gain is left unexploited. Still, if the rise in M in not "too large", the slope of the profit curve at the old price $\bar{P}_i$ may still be small enough to be dominated by the menu cost.

Given a change in M of size $\Delta M > 0$, the opportunity cost of not changing price can be shown to be of "second order", i.e., proportional to $(\Delta M/M)^2$.[15] This is a "very small" number, when $|\Delta M/M|$ is just "small". Therefore, in view of the menu cost, say c, it may be advantageous not to change price. Indeed, the net gain ($= c -$ opportunity cost) by *not* changing price may easily be positive. Suppose this is so for firm i, given that the other firms do not change price. Since each individual firm is in the same situation as long as the other firms have not changed price, the outcome that no firm changes its price is an equilibrium. As in this equilibrium there is no change in the general price level, there will be a higher output level than without the rise in M.

The reference to changes in the money stock, M, in this discussion should not be misunderstood. It is not as a medium of exchange or similar that M has a role in the model, but as the sole constituent of non-human wealth. The increase in M does not reflect an open market purchase of bonds by the central bank, but a money-financed government budget deficit created by transfers to the households without any taxes in the opposite direction. This amounts to a combined monetary-fiscal stimulus to the economy, an example of "helicopter money", cf. Chapter 17.

Doesn't W respond?

The considerations above presuppose that workers or workers' unions do not immediately increase their wage demands in response to the increased demand for labor. This assumption can be rationalized in two different ways. One way is to assume that also the labor market is characterized by monopolistic competition between craft unions, each of which supplies its specific type of labor. If there are menu costs associated with changing the wage claim and they are not too small, the same envelope theorem logic as above applies and so, theoretically, an

[15] Appendix A shows this by taking a second-order Taylor approximation of the opportunity cost.

increase in labor demand need not in the short run have any effect on the wage claims.

There is an alternative way of rationalizing absence of an immediate upward wage pressure. This alternative way is more apt in the present context since we have treated labor as homogeneous, implying that there is no basis for existence of many different craft unions.[16] Instead, let us here assume that *involuntary unemployment* is present. This means that there are people around without a job although they are as qualified as the employed workers and are ready and willing to take a job at the going wage or even a lower wage.[17] Such a state of affairs is in fact what several labor market theories tell us we should expect to see often. In both efficiency wage theory, social norms and fairness theory, insider-outsider theory, and bargaining theory, there is scope for a wage level above the individual reservation wage (see Chapter 24). Presence of involuntary unemployment implies that employment can change with negligible effect on the wage level in the short run. In combination with little price sensitivity to output and employment changes, this observation also offers a rationalization of stylized fact no. 2 in the list of Section 19.1 saying that *relative* prices, including the real wage, exhibit little sensitivity to changes in the corresponding quantities, here employment.

19.3.3 Menu costs in action

Under these conditions even *small* menu costs can be enough to prevent firms from changing their price in response to a change in demand. At the same time even small menu costs can have *sizeable* effects on aggregate output, employment, and social welfare. To understand this latter point, note that under monopolistic competition neither output, employment, or social welfare are maximized in the initial equilibrium. Therefore the envelope theorem does not apply to these variables.

This line of reasoning is illustrated in Fig. 19.5. There are two differences compared with Fig. 19.3. First, aggregate demand is now specified as in (19.38).

[16] Even with heterogenous labor, the craft union explanation runs into an empirical problem in the form of a "too low" wage elasticity of labor supply according to the microeconometric evidence. We come back to this issue in Chapter 20.3.

[17] In case firms have considerable hiring costs (announcing, contracting, and training) these add to the full cost of employing people. Typically there is then an initial try-out period with a comparatively low introductory wage rate. The criterion for being involuntarily unemployed is then whether the person in question is willing to take a job under similar conditions as those who currently got a job.

Although the term "involuntary" may provoke moral sentiment, this definition of *involuntary unemployment* should be understood as purely technical, referring to something that can in principle be measured by observation.

Second, along the vertical axis we have set off the *relative* price, P_i/P, so that marginal revenue, $\mathcal{MR}$, as well as marginal costs, $\mathcal{MC}$, are indicated in real terms, i.e., $\mathcal{MR} = MR/P$ and $\mathcal{MC} = MC/P$. For fixed M/P, the demand curve faced by firm i is shown as the solid downward-sloping curve $D(P_i/P, M/(m\beta P))$ to which corresponds the real marginal revenue curve, $\mathcal{MR}$. For fixed W/P, the real marginal costs faced by the firm are shown as the upward-sloping real marginal cost curve, $\mathcal{MC}$ (recall that we consider the case $\alpha < 1$).

If firms have rational (model consistent) expectations and know M and W in advance, we have $Y^e = M/(\beta P)$. The price chosen by firm i in advance, given this expectation, is then the price $\bar{P}_i$ shown in Fig. 19.3. As the chosen price will be the same for all firms, the relative price, P_i/P, equals 1 for all i. Equilibrium output for every firm will then be $M/(m\beta P)$, as indicated in the figure. If the actual money stock turned out to be higher than expected, say $M' = \lambda M$, $\lambda > 1$, and there were no price and wage adjustment costs *and* if wages were also multiplied by the factor λ, prices would be multiplied by the same factor and the real money stock, production, and employment be unchanged.

With menu costs, however, it is possible that prices and wages do not change. The menu cost may make it advantageous for each single firm not to change price. Then, the higher nominal money stock translates into a higher *real money* stock and the demand curve is shifted to the right, as indicated by the stippled demand curve in Fig. 19.5. As long as $\bar{P}_i/P > \mathcal{MC}$ still holds, each firm is willing to deliver the extra output corresponding to the higher demand. The extra profit obtainable this way is marked as the hatched area in Fig. 19.5. Firms in the other production lines are in the same situation and also willing to raise output. As a result, aggregate employment is on the point of increasing. The only thing that could hold back a higher employment is a concomitant rise in W in response to the higher demand for labor. Assuming presence of involuntary unemployment in the labor market hinders this, the tendency to higher employment is realized, and firm i's production ends up at y_i' in Fig. 19.5, while the price $\bar{P}_i$ is maintained. The other firms act similarly and the final outcome is higher aggregate consumption and higher welfare.

Thus, the effects on aggregate output, employment, and social welfare of not changing price can be substantial; they are of "first order", namely proportional to $|\Delta M/M|$, as implied by the aggregate demand formula (19.38).

In the real world, nominal aggregate demand (here proportional to the money stock) fluctuates up and down around some expected level. Sometimes the welfare effects of menu costs will be positive, sometimes negative. Hence, *on average* the welfare effects tend to cancel out to a first order. This does not affect the basic point of the menu cost theory, however, which is that changes in aggregate nominal demand can have first-order real effects (in the same direction) because

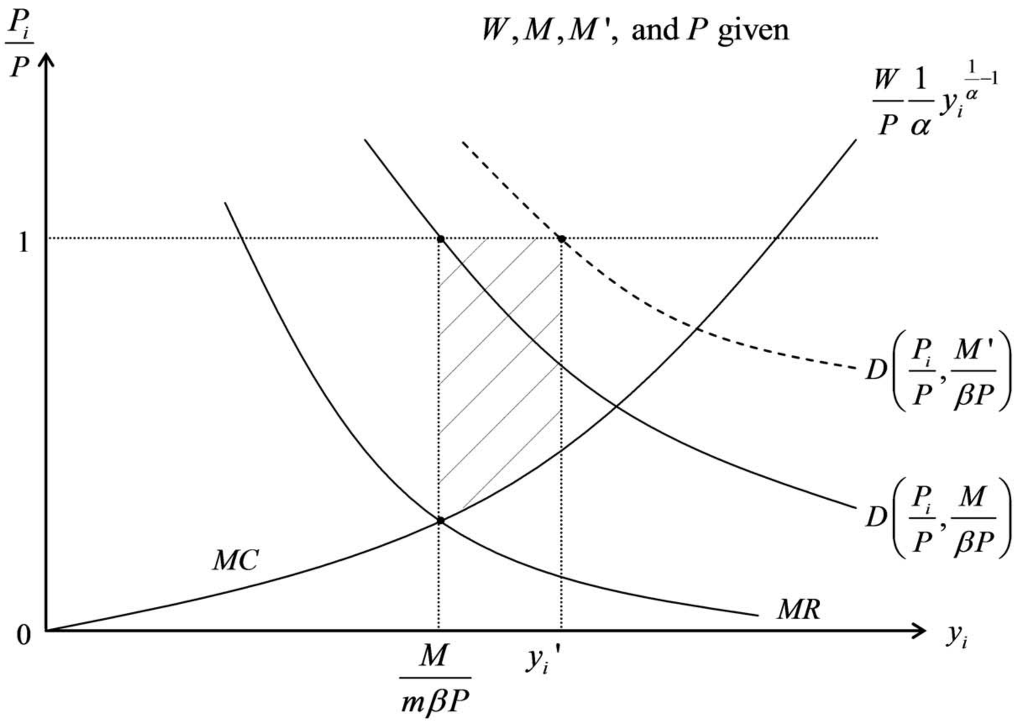

Figure 19.5: The impact in general equilibrium of a shift to $M' > M$ when menu costs are binding (the case $\alpha < 1$).

the opportunity cost by not changing price is only of second order.[18]

A reservation As presented here, the menu-cost story is not entirely convincing. The rather static nature of the setup is a drawback. For instance, the setup gives no clear answer to the question whether it is a change from a given past price, $\bar{P}_{i,t-1}$, to the preset price, $\bar{P}_{i,t}$, for the current period that is costly or a change from the preset price, $\bar{P}_{i,t}$, for the current period to another price within a sub-period of that period.

More importantly, considering a sequence of periods, there would in any period be *some* prices that are not at their ex ante "ideal" level. The firms in question are then not at the flat part of their profit curve. The menu cost necessary to prevent price adjustment will then be higher for these firms, thus making it more demanding for menu costs to be decisive. Moreover, in an intertemporal perspective it is the present value of the expected stream of future gains and costs that matter rather than instantaneous gains and costs. An aspect of a complete dynamic modeling is also that ongoing inflation would have to be taken into account. In modern times where money is paper money (or electronic money), there is usually an underlying upward trend in the general input price level. To maintain profitability, the individual producers will therefore surely *have* to adjust their prices from time to time. The decision about *when* and *how much* to change price will be made with a view to maximizing the *present value* of the expected future cash flow taking the expected menu costs into account.[19]

The key point from static menu-cost theory, based on the envelope theorem, is not necessarily destroyed by the dynamics of price-setting. It becomes less cogent, however.

The rule of the minimum

For a preset price, $\bar{P}_i$, it is beneficial for the firm to satisfy demand as long as the corresponding output level is within the area where nominal marginal cost is below the price. Returning to Fig. 19.3, actual aggregate demand is given as Y^d. Let actual demand faced by firm i be denoted y_i^d, so that $y_i^d = D(\bar{P}_i/P, Y^d/m)$. Compare this demand to y_i^c, defined as the production level at which $MC = \bar{P}_i$ (assuming $\alpha < 1$). This is the production level known as the *Walrasian* or *classical* or *competitive supply* by firm i (the superscript "c" stands for "classical" or "competitive"). It indicates the output that would prevail under perfect competition, given W, $P_i = \bar{P}_i$, and the assumption of rising marginal costs

[18] Sustained increases in aggregate demand are likely to lead to capacity investment by the existing firms or entry of new firms supplying substitutes.

[19] As to contributions within this dynamic perspective, see Literature notes.

($\alpha < 1$). As this desired output level is a function of only W and $\bar{P}_i$, we write it $y_i^c = y^c(W, \bar{P}_i)$. In the case of constant MC, i.e., $\alpha = 1$, we interpret $y^c(W, \bar{P}_i)$ as $+\infty$.

As long as $y_i^d < y_i^c$, and enough labor is available, actual output will be $y_i = y_i^d$. If $y_i^d > y_i^c$, however, the firm will prefer to produce only y_i^c. Producing beyond this level would entail a loss since marginal cost would be above the price. Presupposing enough labor is available, the rule is therefore that given the demand D and the classical supply y_i^c, actual production is the minimum of the two, that is,

$$y_i = \min\left[D(\frac{\bar{P}_i}{P}, \frac{Y^d}{m}), y^c(W, \bar{P}_i)\right].$$

Given the production function $y_i = n_i^\alpha$, $i = 1, 2, \ldots, m$, the corresponding *effective labor demand* by firm i is

$$n_i^d = y_i^{1/\alpha}.$$

The *aggregate effective labor demand* is $N^d = \sum_{i=1}^m n_i^d$. Let the aggregate effective supply of labor be a given constant, $\bar{N}$, and assume that in the short run, $\bar{N}/m$ workers are available to each firm. Then the *effective supply* of firm i is $\min\left[y^c(W, \bar{P}_i), (\bar{N}/m)^\alpha\right]$. Actual output of firm i will be

$$y_i = \min\left[D(\frac{\bar{P}_i}{P}, \frac{Y^d}{m}), y^c(W, \bar{P}_i), (\frac{\bar{N}}{m})^\alpha\right], \tag{19.42}$$

that is, the minimum of effective demand, classical supply, and output at full employment in "product line i". This rule is known as the *rule of the minimum.*[20]

If at the given wage and price level, labor supply is the binding constraint in most of the product lines, *repressed inflation*, as defined in Section 19.2, prevails with *excess demand* for labor and goods.

Keynesian versus classical unemployment In the opposite case, where labor is abundant, two alternative kinds of unemployment may prevail. If the demand, D, is the binding constraint in most of the product lines, what is known as *Keynesian unemployment* prevails. This is a situation where both the typical output market and the labor market are in a state of "buyers' market" (sale less than preferred).

[20]Note that this rule determines production of the single firm. It is related to, but not identical to the *short-side rule,* which we encountered in Section 19.2. This is the "voluntary trade" principle saying that the actual quantity traded in a market is the minimum of effective demand and effective supply in the market.

The alternative possibility is that the classical supply, y^c, is the binding constraint in most of the product lines. In this case what is known as *classical unemployment* will prevail. This is a situation with downward pressure on the wage level and upward pressure on the price level. The huge unemployment during the Great Depression in the interwar period was by most economists of the time diagnosed as a momentary phenomenon caused by a "too high real wage". Keynes and a few like-minded disagreed. It is in this context that the quote by the outstanding French economist Edmond Malinvaud (1923-) at the front page of this chapter should be seen.

In the *World's Smallest Macroeconomic Model* of Section 19.2 classical unemployment can not occur because of constant marginal costs combined with a positive mark-up μ. The limiting case $\alpha = 1$ in our present disaggregate model also leads to $MC = W$, a constant. Thus (19.42) gives $y_i = D(\bar{P}_i/P, Y^d/m) < y^c(W, \bar{P}_i) = +\infty$ for all i. All the goods markets are demand-constrained and any unemployment is thereby Keynesian. Note also that because of constant MC combined with the constant mark-up, no menu costs are needed to maintain that the output level rather than prices respond to changes in aggregate demand, Y^d.

Although constant MC within certain limits may be an acceptable assumption, an additional factor potentially constraining production in the short run is the capital equipment of the firm. Hitherto this factor has not been visible. Or we might say that the case of rising MC ($\alpha < 1$) can be interpreted as reflecting that in practice labor is not the only production factor. This motivates the next section.

19.4 Abundant capacity

One of the stylized facts listed in Section 19.1 is that under "normal circumstances" a majority of firms in an industrialized economy respond to short-run shifts in aggregate demand by adjusting production rather than price. Key elements in the explanation of this phenomenon have been sketched: (a) the distinction between *Walrasian* and *effective* demand and supply; (b) price setting agents in markets with imperfect competition; (c) the "envelope argument" that the potential benefit of adjusting the price can easily be smaller than the cost of adjusting; and (d) because prices are generally above marginal costs, firms are willing to adjust production when aggregate demand shocks occur.

This leads us to the problem whether quantity adjustment is in the main *realizable* in the short run. Under the assumption that involuntary unemployment is present, lack of workers will not be an impediment. But the production capacity of firms depends also on their command of capital equipment. To throw light on this aspect we now let the production function have two inputs, capital and labor.

19.4.1 Putty-clay technology

Suppose firm i has the production function

$$y_i = f(k_i, n_i), \tag{19.43}$$

where k_i is the installed capital stock and n_i the labor input, $i = 1, 2, \ldots, m$. (At the disaggregate level we use small letters for the variables. So, contrary to earlier chapters, k_i is here not the capital-labor ratio, but simply the capital stock in firm i.) Because of strictly convex installation costs, k_i is given in the short run. Raw materials and energy are ignored. So in the short run the capital costs are fixed but labor costs variable since n_i can be varied.

Realistic short-run analysis makes a distinction between the "ex ante" and the "ex post" production function. By *ex ante* is here meant the point in time where the decision about investment, whether in plant or equipment, is to be made. We imagine that in making this decision, a wide range of production techniques (input-output combinations) is available as represented by the function f in (19.43), hence called the ex ante production function. The decision will be made with a forward-looking perspective. Construction and installation are time consuming and to some extent irreversible.

By *ex post* is meant "when construction and installation are finished and the capital is ready for use". In this situation, the substitutability between capital and labor tends to be limited. Our long-run models in previous chapters implicitly ignored this aspect by assuming that substitutability between capital and labor is the same ex ante and ex post. In reality, however, when a machine has been designed and installed, its functioning will often require a more or less fixed number of machine operators. What can be varied is just the *degree of utilization* of the machine per time unit. In the terminology of Section 2.5 of Chapter 2, technologies tend in a short-run perspective to be "putty-clay" rather than "putty-putty".

An example: suppose the production function f in (19.43) is a neoclassical production function. This is our ex ante production function. Ex post, this function no longer describes the choice opportunities for firm i. These are instead given by a Leontief production function with CRS:

$$y_i = \min(Au_i\bar{k}_i, Bn_i), \qquad A > 0, B > 0, \tag{19.44}$$

where A and B are given technical coefficients, $\bar{k}_i$ is the size of the installed capital (now a fixed factor) and u_i its utilization rate ($0 \leq u_i \leq 1$), $i = 1, \ldots, m$.[21] Presumably the firms would have acquired their capital equipment under different

[21] The link between the ex ante production function f and the technical coefficients A and B was described in Chapter 2.

circumstances at different points in time in the past so that, generally, the equipment would be somewhat heterogeneous and A and B would be index numbers and differ across the firms. To make aggregation simple, however, the analyst may be tempted in a first approach to ignore this complication and assume A and B are the same across the firms.

19.4.2 Capacity utilization and monopolistic-competitive equilibrium

Capacity utilization in a Keynesian equilibrium

There is *full capacity utilization* when $u_i = 1$, which means that each machine is operating "full time" (seven days and nights a week, allowing for surplus time for repairs and maintenance). Capacity is given as $A\bar{k}_i$ per week. Producing efficiently at capacity requires $n_i = A\bar{k}_i/B$. But if demand, y_i^d, is less than capacity, satisfying this demand efficiently requires $n_i = y_i^d/B$ and $u_i = Bn_i/(A\bar{k}_i) < 1$. As long as $u_i < 1$, there is unused capacity, and marginal productivity of labor in firm i is a *constant*, B.

The (pure) profit of firm i is $\Pi_i = P_i y_i - \mathcal{C}(y_i, W, F_i)$, where $\mathcal{C}(y_i) = Wy_i/B + F_i$ is the cost function with F_i denoting the fixed costs deriving from the fixed production factor, $\bar{k}_i$. Average cost is $AC = \mathcal{C}(y_i)/y_i = W/B + F_i/y_i$ and marginal cost is a constant $MC = W/B$ for $y_i < A\bar{k}_i$.

Fig. 19.6 depicts these cost curves together with a downward-sloping demand curve. A monopolistic-competition market structure as described in Section 19.3.1 is assumed but now only a subset of the m firms produce differentiated consumption goods. The other firms produce differentiated capital goods, also under conditions of monopolistic competition and with the same price elasticity of demand.

Firm i presets the price of good i at $P_i = \bar{P}_i$ in the expectation that the level of demand will be as indicated by the downward sloping D curve in Fig. 19.6. The point E_{SR} in the figure represents a standard short-run equilibrium under monopolistic competition with output level such that $MC = MR$. Assuming full symmetry across the different firms, the point E_{SR} would also reflect a Keynesian equilibrium if the actual demand level (position of the demand curve) had turned out to be as expected by the firms when fixing their price. But in the figure it is assumed that the demand level turned out lower. The produced quantity is reduced while the price remains unchanged (because of either menu costs or simply the constant marginal costs combined with constant price elasticity of demand). Firm i ends up with actual production equal to y_i' in Fig. 19.6. The obtained (pure) profit is indicated by the hatched rectangle constructed by the help of the average cost curve AC.

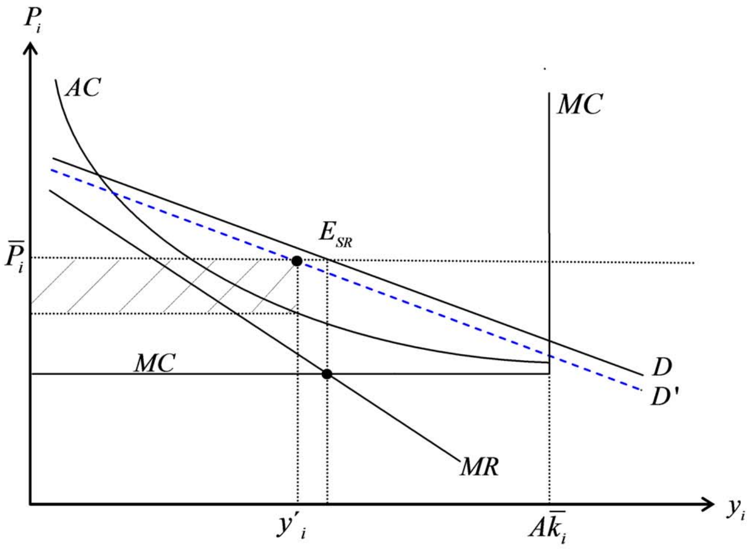

Figure 19.6: Firm i in a Keynesian equilibrium ($A\bar{k}_i$ is production capacity; MC curve first horizontal and then vertical at $y_i = A\bar{k}_i$).

By interpreting y_i' in Fig. 19.6 as actual production we have implicitly assumed that enough labor is available. We let this "labor abundance" be understood throughout this discussion. If the picture in Fig. 19.6 is representative for the economy as a whole, the unemployment in the economy is predominantly *Keynesian.*

If the actual demand level had turned up higher than expected, firm i would be induced to raise production. There is scope for this because price is above marginal costs the whole way up to full capacity utilization. Profits will by this expansion of production become *substantially* higher than expected because the profit on the marginal unit sold is higher than average profit per unit as a result of the AC curve being downward-sloping up to the production level $A\bar{k}_i$.

Anyway, all the way up to $A\bar{k}_i$ we have a situation where the quantity produced is less than the quantity at which average costs are at the minimum. Such a situation is sometimes said to reflect "excess capacity". But "excess" sounds as if the situation reveals a kind of inefficiency, which need not be the case. So we prefer the term "abundant capacity".

A monopolistic-competitive long-run equilibrium

The picture is essentially the same in a "free-entry-and-exit equilibrium" where all pure profit is eliminated; in the theory of industrial organization this is known as a "long-run equilibrium". Suppose that initially some of the firms get positive

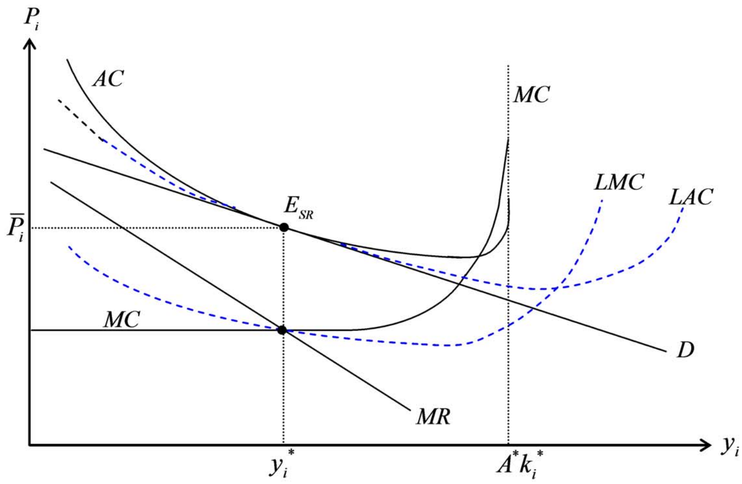

Figure 19.7: Firm i in a monopolistic-competition long-run equilibrium (y_i^* is long-run equilibrium output and $A^*k_i^*$ is long-run production capacity).

pure profits as illustrated in Fig. 19.6. This state of affairs invites entry of new firms. Over time these set up new plants and begin to supply new differentiated goods from a large set of as yet un-utilized *possible* imperfect substitutes for the existing goods. The entry process continues until equilibrium with zero pure profits applies to all product lines. When this state is reached, prices equal both short- and long-run average costs, and each firm operates where the downward-sloping demand curve is *tangent* to both AC curves. By "short-run" we mean a time horizon within which only a subset of the production factors are variable, while "long-run" refers to a time horizon long enough for all production factors to be variable.

From a macroeconomic perspective the important conclusion is that the "transition" from short-run equilibrium to long-run equilibrium in no way tends to lessen the presence of abundant capacity in the firms.

To portray a long-run equilibrium, we only need to let the AC curve for firm i's chosen plant and equipment be tangent to the demand curve at the point E_{LR}. This is what we have done in Fig. 19.7 where also the long-run marginal and average cost curves are visible, denoted LMC and LAC, respectively. The LMC curve is assumed U-shaped. This implies a U-shaped LAC curve. The downward-sloping part of the LMC curve may be due to indivisibilities of plant and equipment. And the upward-sloping part may reflect coordination problems or an implicit production factor which is tacitly held fixed (a special managerial expertise, say).

Independently of the long-run versus short-run perspective, we have in Fig. 19.7 introduced the case where the short-run marginal cost curve, MC, is horizontal only up to certain rate of capacity utilization $\bar{u} < 1$. It then rises gradually and ends up as vertical at full capacity utilization. This is to open up for the possibility that a decision to produce more here and now may imply bringing less efficient standby equipment to use. Also the wear and tear on the machinery may be raised. This amounts to a "rounding off" of the possibly too "sharp" Leontief production function (19.44). Instead, efficient production could here be described by

$$n_i = \begin{cases} y_i/B & \text{if } y_i < \bar{y}_i \equiv \bar{u}A\bar{k}_i, \\ \bar{y}_i/B + (y_i - \bar{y}_i)^{1/\alpha}/B & \text{if } \bar{y}_i \leq y_i \leq A\bar{k}_i, \quad 0 < \alpha < 1. \end{cases}$$

An alternative or additional reason for the MC curve to be upward-sloping at high capacity utilization is wage bonuses for working on the night shift or in the weekend.

Also in this more realistic setup is *abundant capacity* revealed as a sustainable equilibrium phenomenon. In long-run equilibrium each firm produces at a point where:

- price is above marginal costs so as to exactly cover fixed costs;
- the quantity produced is less than the quantity at which average costs are at the minimum (i.e., where the MC curve crosses the AC curve in Fig. 19.7), given the firm's preferred plant and equipment.

The conclusion is that as long as the AC curve does not shift (this would happen if the general wage level changed), firms are more than willing to accommodate an increased demand (outward shift of the demand curve) at an unchanged price or even at a lower price by an increase in production. Similarly, an inward shift of the demand curve will not lead to a temptation to reduce the price, rather the opposite if the menu cost is immaterial. These observations fit well with the huge amount of sales promotion we see. They also fit with the empirical evidence that measured total factor productivity and gross operating profits rise in an economic upturn and fall in a downturn, an issue to which we return in Part VII of this book.

Finer shades*

Oligopoly In the real world some markets are better characterized by strategic interaction between a few big firms than by monopolistic competition. This is a situation where abundant capacity may result not only from falling average costs but also from a strategic incentive. Maintaining abundant capacity will make credible a threat to cut price in response to unwelcome entry by a competitor.

Contestable markets A *contestable market* is a market for a homogeneous good where, because of large economies of scale, the quantity at which average costs are at the minimum exceeds the size of the market so that there is only room for one firm if production costs should be minimized.[22] The mere "threat of entry" induces average cost-pricing by the incumbent. So, with the point E_{LR} interpreted as the long-run equilibrium under these conditions, Fig. 19.7 also portrays this situation. Again "abundant capacity" is displayed.

The role of indivisibilities The downward-sloping part of the LAC curve, reflecting indivisibilities in plant and equipment, is important also from another perspective. Without indivisibilities it may be difficult to see why involuntarily unemployed workers could not just employ and support themselves in the backyard, financing the needed tiny bits of capital by tiny bank loans. This point is developed further in, e.g., Weitzman (1982).

19.4.3 Aggregation over different regimes*

Returning to Figure 19.6, let us assume that the position of the demand curve faced by firm i is shifted to the right so that the production capacity $A\bar{k}_i$ becomes a binding constraint on y_i'. Suppose further that most of the industries are in this situation. If unemployment is still massive, it is no longer mainly Keynesian, but a particular form of *classical* unemployment. Neither aggregate demand nor production costs as such like in Section 19.3.1, but simply the lack of sufficient capital is the binding constraint on employment. This state of affairs amounts to a form of classical unemployment known as *Marxian unemployment* because it was emphasized in the economic writings of Karl Marx.

Even though the phenomenon of insufficient capital is generally regarded as more common in developing countries (and in the pre-industrial period of Western Europe), it also appears from time to time in specific product lines of industrialized countries under structural change. Similarly, the constraint from labor supply may from time to time be binding in other product lines. Macroeconomic analysis should therefore allow for different regimes in the different product lines of the economy.

Let us imagine that firm i (or industry i) not only produces its own differentiated good i but is also distinguished by using a particular type of labor, say "local labor", effectively supplied in the amount n_i^s. According to the rule of the minimum, actual production will then be

$$y_i = \min(y_i^d, y_i^c, y_i^f), \qquad i = 1, \ldots, m,$$

[22] Cf. Tirole (1988).

where $y_i^c \equiv A\bar{k}_i$ and $y_i^f \equiv Bn_i^s$. Depending on which is the binding constraint, y_i^d, y_i^c, or y_i^f, firm i is either in the Keynesian regime, the classical regime, or the repressed-inflation regime.

Because of technological change and changes in demand patterns we expect some regime heterogeneity, known as *mismatch*, to evolve in the economy as a whole. For the aggregate level, we define $YD \equiv \sum_i y_i^d$, $YC \equiv \sum_i y_i^c$, and $YF \equiv \sum_i y_i^f$. Then, in general, $Y \equiv \sum_i y_i < \min(YD, YC, YN)$.

As an alternative to an aggregate *min* condition as in Section 19.2, a large multi-country study of Western European unemployment since the 1960s, entitled *Europe's Employment Problem*,[23] introduced a statistical distribution of demands and supplies on micro-markets for goods and labor in each country at a given point in time. The approach can briefly be described as follows. Suppose the micro-market values y^d and y^c are jointly log-normally distributed:

$$\begin{bmatrix} \log y^d \\ \log y^c \end{bmatrix} \sim \mathcal{N}\left[\begin{pmatrix} \xi_d \\ \xi_c \end{pmatrix}, \begin{pmatrix} \sigma_d^2 & \text{cov} \\ \text{cov} & \sigma_c^2 \end{pmatrix}\right].$$

Then $\log(y^d/y^c) \sim \mathcal{N}(\xi_d - \xi_c, \sigma^2)$, where $\sigma^2 = \sigma_d^2 + \sigma_c^2 - 2\text{cov}$. Letting Y^* denote firms' *aggregate desired output* (aggregate output in case labor supply nowhere is the binding constraint), it can be shown[24] that

$$Y^* \approx (YD^{-\rho} + YC^{-\rho})^{-1/\rho}, \qquad \rho > 0, \tag{19.45}$$

that is, Y^* is approximately a constant-returns-to-scale CES function of YD and YC. The inverse of ρ is then an increasing function of the variance of $\log(y^d/y^c)$ and is therefore a measure of the "degree of mismatch" between the demand constraint and the capacity constraint. Indeed, it can be shown that $(YD^{-\rho} + YC^{-\rho})^{-1/\rho} < \min(YD, YC)$ for $\rho \in (0, \infty)$ and that $\lim_{\rho\to\infty}(YD^{-\rho} + YC^{-\rho})^{-1/\rho} = \min(YD, YC)$, saying that as $1/\rho \to 0$, mismatch on the firms' side disappears.

As to mismatch in the labor markets, consider firm i's actual employment $n_i = \min(n_i^d, n_i^s)$, where $n_i^d = \min(y_i^d, y_i^c)$ is effective demand for labor, and n_i^s is the effective supply of labor. Given the Leontief production function (19.44), the aggregate effective demand for labor is

$$N(Y^*) = \frac{Y^*}{B} = \left((\frac{YD}{B})^{-\rho} + (\frac{YC}{B})^{-\rho}\right)^{-1/\rho} \equiv (NY^{-\rho} + NC^{-\rho})^{-1/\rho}. \tag{19.46}$$

In analogy with (19.45) we assume that actual aggregate employment satisfies

$$N \approx (N(Y^*)^{-\rho'} + NS^{-\rho'})^{-1/\rho'}, \qquad \rho' > 0,$$

[23] See Drèze and Bean (1990).

[24] For the math behind this and other claims in this section, see Lambert (1988).

where $NS \equiv \sum n_i^s$ and ρ' measures the "degree of mismatch" between the demand and supply in the labor markets. Substituting (19.46) into this gives

$$N \approx \left((NY^{-\rho} + NC^{-\rho})^{-\rho'/\rho} + NS^{-\rho'}\right)^{-1/\rho'} = \left(NY^{-\rho} + NC^{-\rho} + NS^{-\rho}\right)^{-1/\rho},$$

where the last equality holds *if* $\rho' = \rho$. In that case we also have

$$Y = BN = (YD^{-\rho} + YC^{-\rho} + YS^{-\rho})^{-1/\rho}, \tag{19.47}$$

approximately, where $YS \equiv B \cdot NS$ and where, for convenience, we have replaced "$\approx$" by "$=$", appealing to the law of large numbers.

The parameters ρ and ρ' can be estimated on the basis of business and household survey data (firms' answers to regular survey questions about demand, capacity and labor constraints and households' answers about desired employment). The mentioned *Europe's Employment Problem* study estimated for most of the countries (Denmark included) a falling ρ since middle of the 1960s towards the late 1980s, that is, a rising mismatch. This tends to raise the unemployment rate. To illustrate, imagine the "favorable" case where $NY = NC = NS$ so that without mismatch full-employment equilibrium would prevail. Actual employment will be

$$N = (3NS^{-\rho})^{-1/\rho} = 3^{-1/\rho} NS.$$

The unemployment rate then is

$$u \equiv \frac{NS - N}{NS} = 1 - \frac{N}{NS} = 1 - 3^{-1/\rho} > 0,$$

when $\rho < \infty$, i.e., $1/\rho > 0$. An increased mismatch, $1/\rho$, thus means a higher u.

Another consequence of mismatch is that it reduces the Keynesian spending multiplier. Consider aggregate demand as given by the standard textbook income-expenditure equation

$$YD = C(Y) + \bar{I} + \bar{G} + \bar{X} - IM(Y), \qquad C' > 0,\ IM' > 0, 0 < C' - IM' < 1, \tag{19.48}$$

where $C(Y)$ and $IM(Y)$ are private consumption and imports, respectively, while $\bar{I}$, $\bar{G}$, and $\bar{X}$ are private investment, government purchases, and exports, respectively, all exogenous. To find the multiplier with respect to $\bar{G}$, we substitute (19.48) into (19.47) and differentiate with respect to $\bar{G}$, using the chain rule:

$$\frac{\partial Y}{\partial \bar{G}} = -\frac{1}{\rho}(YD^{-\rho} + YC^{-\rho} + YS^{-\rho})^{-\frac{1}{\rho}-1} \cdot (-\rho) YD^{-\rho-1} \left((C' - IM')\frac{\partial Y}{\partial \bar{G}} + 1\right).$$

By ordering,

$$\frac{\partial Y}{\partial \bar{G}} = \frac{1}{\left(\frac{YD}{Y}\right)^{\rho+1} - C' + M'} < \frac{1}{1 - C' + M'}, \tag{19.49}$$

where the inequality is due to $Y < YD$. In turn, this latter inequality reflects that not all micro-markets are in a Keynesian regime.[25]

The mentioned multi-country study thus concluded that increased mismatch in the preceding years was part of the explanation of the high level of unemployment in Western Europe in the 1980s. Moreover, the increased mismatch was attributed to the collapse of capital investment in the aftermath of the first and second oil price crises 1973 and 1979. As a consequence a higher fraction of industries, $y_i^c - y_i^d$ became negative.

Additional conclusions for the period considered, from the middle of the 1960s to the late 1980s, were:

1. Keynesian unemployment has been the dominant regime.
2. The influence of demand pressure on prices has been negligible; instead demand pressures spill over into increased imports.
3. The degree of capacity utilization has been a significant determinant of investment.
4. The elasticity of prices with respect to wage costs is substantial, ranging from 0.5 in the short run to 1.0 in the long run.
5. Increases in real wages induce capital-labor substitution.
6. The main determinant of output growth in the eighties in Western Europe has been effective demand.

19.5 Concluding remarks

(incomplete)

Let us summarize. This chapter has extended the *income-expenditure model*, known from introductory textbooks, with some microeconomic underpinnings. The framework is based on the idea that for short-run analysis of effects of demand shocks in an industrialized economy it makes sense, as a first approximation, to treat the nominal price level as a predetermined variable. We have built on the

[25] Warning: As $YD/Y > 1$, from (19.49) it may appear that a rise in mismatch, i.e., decrease in ρ, raises the Keynesian multiplier $\partial Y/\partial \bar{G}$. This counter-intuitive impression is false, however. Indeed, a decrease in ρ means an increased YD/Y through a reduced Y, cf. (19.47).

assumption that the wage level is predetermined and that the marginal cost curve is horizontal. When this is combined with constant markups due to a more or less constant price elasticity of demand at firm level, a price level which is independent of output in the short run follows.

An alternative - or supplementary - approach to the explanation of the presumed price insensitivity, building on the menu cost theory, was also described. The idea here is that there are fixed costs (pecuniary or non-pecuniary) associated with changing prices. The main theoretical insight of the menu cost theory is that even *small* menu costs can be enough to prevent firms from changing their price. This is because the opportunity cost of not changing price is only of second order, i.e., "small"; this is a reflection of the envelope theorem. So, nominal prices may be sticky in the short run even if marginal costs, MC, are rising. But owing to imperfect competition (price $> MC$), the effect on aggregate output, employment, and welfare of not changing prices is of first order, i.e., "large".

The described framework allows us to think in terms of general equilibrium, in the sense of a state of rest, in spite of the presence of some non-clearing markets. First and foremost the labor market belongs to the latter category. A key distinction is the one between *effective* supplies and demands and *actual* transactions.

Apart from the border case where all markets clear, three different types of short-run equilibria arise: repressed inflation, classical unemployment, and Keynesian unemployment. The Keynesian view is that the latter type of short-run equilibrium is prevalent in industrialized economies. Repressed inflation seems rare. We may put it this way: wages and prices appear less sticky in situations with upward pressure than in situations with downward pressure. Moreover, as long as there is a positive mark-up, classical unemployment is ruled out, at the theoretical level, if constant short-run marginal costs are assumed. Not all macroeconomists regard such an assumption empirically tenable, especially with a view on peak periods in the business cycle.

Abundant capacity. Micro-markets. Mismatch.

Empirics on price stickiness: Blinder ( ).

Bils, Klenow, and Malin (2012) find evidence in support of Keynesian labor demand

A rigorous general equilibrium model with monopolistic competition, the Blanchard-Kiyotaki model,[26] is set up and analyzed in the next chapter. That model includes a complete description of the households with respect to preferences regarding differentiated consumption goods and supply of different types of labor. Still only a single financial asset is available, base money. Readers eager to attribute to asset markets a more important role may jump directly to

[26] Blanchard and Kiyotaki (1987).

Chapter 21. That chapter presents and analyzes the IS-LM model, based on John R. Hicks' summary of the analytical content of Keynes' main opus, *The General Theory of Employment, Interest and Money* from 1936.[27] This summary became a cornerstone of mainstream short-run macroeconomics after the Second World War.

Throughout this chapter the functioning of labour markets has received scant attention. As alluded to, the Blanchard-Kiyotaki model of the next chapter represents one approach to the integration of labor markets. In Chapter 24 other approaches are discussed.

19.6 Literature notes

(incomplete)

The basic model in Keynes' *General Theory* (1936) relied less on imperfect competition than what became normal in later Keynesian thinking, as articulated for instance by the World's Smallest Macroeconomic Model and the various new-Keynesian contributions to be considered later. In Keynes (1936) only the labor market has imperfect competition, resulting in a predetermined nominal wage level. In the output market, perfect competition with full price flexibility rules. This feature, including Keynes' associated conjecture that real wages would be countercyclical, was criticized on empirical grounds by Dunlop (1938) and Tarshis (1939). In his answer, Keynes (1939) acknowledged the need for reconsideration of this matter.

In the vocabulary of Walrasian economics, the term *equilibrium* is reserved to states where all markets clear unless the price in question has fallen to zero. On this background it may be a surprise that one may talk about Keynesian *equilibrium* with unemployment. But equilibrium is an abstract concept that need not require equality of *Walrasian* demand and supply. Walrasian equilibrium is just *one* type of equilibrium, *one* type of state of rest for an economic system. In this chapter we have introduced *another* kind of state of rest, relevant under other circumstances: *equilibrium with quantity rationing.* By adhering to this terminology, we follow the strand within macroeconomics called "macroeconomics with quantity rationing", to which the French scholar Edmund Malinvaud, cited in the introduction to this chapter, belongs. The first to show existence of general equilibrium in a fully articulated disaggregate setup, but with fixed prices and quantity rationing, was another French economist, Jean-Pascal Benassy (1975). An important precursor is Chapter 14 in Arrow and Hahn (1971).

There is an alternative terminology in which a state of affairs with non-clearing

[27] Hicks,1937.

markets, in the Walrasian sense, is termed a *disequilibrium*. The title, "A general disequilibrium model of income and employment", of the seminal paper by the American economists Robert Barro and Herchel Grossman (1971) is a case in point as is the vocabulary in the book *On Keynesian Economics and the Economics of Keynes* by Axel Leijonhufvud (1968).[28] Although terminologies differ, contributors to these strands of macroeconomics, including Patinkin (1956), Clower (1965), and Herings (1996), seem to agree that the important aspects are the dynamic processes triggered by non-clearing markets. A thorough account of macroeconomics with quantity rationing is given in Malinvaud (1998b). The theory has been applied to analytical studies of mass unemployment as in, e.g., Malinvaud (1984, 1994) and empirical studies like the large econometric multi-country study entitled *Europe's Employment Problem* (1990), based on the theoretical framework in Sneessens and Drèze (1986) and Lambert (1988).

Already Karl Marx (1867) rejected Say's law by emphasizing the option of hoarding money instead of buying produced goods. A contemporary examination of the role of Walras' law and refutation of Say's law in macroeconomics is contained in Patinkin (2008). Recurring controversy about Say's "Law", or Say's "fallacy", as some opponents say, arise during periods of severe recession or depression. During the Great Depression Keynes charged the UK Treasury and contemporary economists for being "'deeply steeped in the notion that if people do not spend their money in one way they will spend it in another" (Keynes, 1936, p. 20). By a series of citations, DeLong (2009) finds that a similar notion characterizes one side in "the fiscal stimulus controversy" in the US in the aftermath of the financial crisis 2007-08.

Keynes (1937).

Comparison between Keynes (1936) and Keynes (1939). Keynes and Hicks. Kalecki.

In Section 19.2 we assumed that he perceived quantity signals, like the price signals, are deterministic. In Svensson.( ) the theory is extended to include stochastic quantity signals.

Market forms: For estimations of the markup in various U.S. industries, see, e.g., Hall (1988).

Even a dynamic general equilibrium with perfect competition is not a completely lucid thing. In perfect competition all firms are price takers. So who is left to change prices? This is a sign of a logical difficulty within standard competitive theory (as pointed out by Arrow, 1959). Merely assigning price setting to abstract "market forces" is not theoretically satisfactory, and reference to the

[28] A few years after the publication of this paper, Robert Barro lost confidence in the Keynesian stuff and became one of the leading new-Classical macroeconomists. His reasons are given in Barro (1979).

mythical "Walrasian auctioneer" is not convincing.

Reference to inventory dynamics?

The convex price adjustment cost approach, Rotemberg, JME, 52 (4), 1982, 829-852.

Approaches to *menu costs in a dynamic context*:

Caplin and Spulper (1987), briefly summarized in Benassy (2011, p. 317-18).

The inclusion of ongoing inflation, see Blanchard 1990 and Jeanne 1998.

One might conjecture Ginsburg et al., EL, 1991.

Danziger (AER PP, 1999, SJE 2008).

Bursteain and Hellwig (AER, 2008).

Caballero and Engel (2007).

What about costs of changing the production level?

19.7 Appendix

A. The envelope theorem

ENVELOPE THEOREM FOR AN UNCONSTRAINED MAXIMUM Let $y = f(a,x)$ be a continuously differentiable function of two variables, of which one, a, is conceived as a parameter and the other, x, as a control variable. Let $g(a)$ be a value of x at which $\frac{\partial f}{\partial x}(a,x) = 0$, i.e., $\frac{\partial f}{\partial x}(a,g(a)) = 0$. Let $F(a) \equiv f(a,g(a))$. Provided $F(a)$ is differentiable,

$$F'(a) = \frac{\partial f}{\partial a}(a,g(a)),$$

where $\partial f/\partial a$ denotes the partial derivative of $f(a,x)$ w.r.t. the first argument.

Proof $F'(a) = \frac{\partial f}{\partial a}(a,g(a)) + \frac{\partial f}{\partial x}(a,g(a))g'(a) = \frac{\partial f}{\partial a}(a,g(a))$, since $\frac{\partial f}{\partial x}(a,g(a)) = 0$ by definition of $g(a)$. □

That is, when calculating the total derivative of a function w.r.t. a parameter and evaluating this derivative at an interior maximum w.r.t. a control variable, the envelope theorem allows us to ignore the terms that arise from the chain rule. This is also the case if we calculate the total derivative at an interior minimum.[29]

B. The opportunity cost of not changing price is of second order

(no text yet available)

[29] For extensions and more rigorous framing of the envelope theorem, see for example Sydsaeter et al. (2006).

19.8 Exercises