

**Swedish Institute for Social Research (SOFI)**

---

**Stockholm University**

---

**WORKING PAPER 1/2018**

**BASIC MACROECONOMICS**

**by**

**Ante Farm**

# BASIC MACROECONOMICS

Ante Farm  
ante.farm@sofi.su.se  
Swedish Institute for Social Research (SOFI), Stockholm University  
SE-106 91 Stockholm, Sweden  
www.sofi.su.se

January 5, 2018

*Abstract:* While traditional macroeconomics is organized around different theories for the short-run, the medium run and the long run, this text is an introduction to macroeconomics which applies to every year, not only every year of the short run, but also every year of the medium run and the long run. Based on a new benchmark model of pricing, it clarifies the determinants of inflation. Based on macroeconomic identities, it clarifies the determinants of aggregate profits. Based on a complete model of labour demand, it clarifies the determinants of employment. And based on the definition of unemployed as people without employment looking for jobs, it clarifies the determinants of unemployment, including matching problems.

*Keywords:* Output, inflation, employment, unemployment.

*JEL classification:* E23, E24, E31

*Acknowledgements:* Detailed comments on earlier drafts by Nils Lindman, Lars-Erik Lögdberg, and Stefan Norrman are gratefully acknowledged. I am also grateful for comments, references and advice from Lennart Erixon, Harry Flam, Erik Hegelund, Tomas Korpi, Per Lundborg, and Eva Skult.

## Introduction

### What is *basic* macroeconomics?

Traditional macroeconomics is organized around three different theories: one for the short run (year-to-year changes), one for the medium run (over a decade or so), and one for the long run (many decades). The long run deals with *growth* of output (as measured by GDP at constant prices). The intermediate run focuses on *levels* of inflation and unemployment, while the short run is concerned with *fluctuations* around these levels.

More precisely, short-run models deal with fluctuations of output, inflation, and unemployment around “equilibrium” values. Fluctuations are composed of deviations from equilibrium caused by “shocks” and convergence towards equilibrium caused by monetary policy and (sometimes) fiscal policy. Shocks are (sudden) changes in aggregate demand (including government spending), oil prices, wages, etc. An analysis of economic fluctuations can focus on either output or unemployment, assuming that these variables are related by Okun’s law. And monetary policy affects output and inflation by exploiting two relations, first the IS curve, which shows how output depends on the policy rate set by the central bank and, second, the Phillips curve, which shows how inflation depends on unemployment or output.

The current *level* of a macroeconomic variable is usually called its *equilibrium* value, but a more neutral term is *normal* value. For example, normal unemployment can be defined as “the rate that prevails when the economy is neither in a boom nor a recession”, as suggested by Jones (2014 p. 184), or, more precisely, as the *average* over a relevant period.

Now, a special theory for the intermediate run of a macroeconomic variable presupposes that its level (average value) during a period of several years is determined before, or apart from, its yearly development. On the other hand, the values of a macroeconomic variable depend on its development, so that, for example, employment in the long run – or more precisely the development of employment over time – is determined by the history of its growth, that is, by the accumulation over time of yearly changes.

How can these two approaches be reconciled? How can the level (normal value) of a macroeconomic variable during a certain period on one hand be “exogenously” determined, and on the other hand be determined “endogenously” by its development during the period? The simplest answer is that the level is indeed exogenously determined, like a target for inflation. Thus, in an economy with inflation targeting it should come as no surprise that the level of inflation is equal to this target, at least approximately, as emphasized by Jones (2014

p. 232). On the other hand, the level of inflation during a certain period is *ultimately* determined endogenously by its development during this period, and this development can imply not only fluctuations around the target but also, for example, a level below the target – as recently observed in many countries. And to understand such a development we must study the determinants of inflation each year.

In long-run or medium-run theories of economic growth or unemployment it is more difficult to assume that the level – the normal value during a certain period – is exogenously determined by economic policy. On the other hand, economic policy will probably affect the level of both unemployment and economic growth, both directly (e.g. by targeting unemployment) or indirectly (e.g. through industrial policy or trade policy), even if other factors may dominate. *In all cases, however, the level (average value) of a macroeconomic variable during a period is ultimately determined by its values in each year of the period.* And in this text I focus on the determinants of inflation, output, employment, and unemployment during a year. This is *basic* macroeconomics.

This approach also suggests an *evolutionary approach* to economics. Thus, the state of an economy is ultimately explained by its *history*, that is, the cumulative result of yearly changes, with a more or less distant past as an unexplained point of departure.

Moreover, focusing on the functioning of the market mechanism during a year, this text also attempts to clarify the determinants of variables like inflation and unemployment without recourse to systems of equations (like the IS-LM model, the AS-AD model, or the 3-equation model) and the corresponding graphs. And this approach is based on the assumption that a market economy develops recursively during a year, starting with changes in wages, followed by changes in prices, followed by changes in sales, production and employment, followed by changes in unemployment.

But before we can discuss wage formation we first have to clarify the effects of wages on prices, employment and unemployment, since wage-setting agents must be able to anticipate the effects of wages on crucial macroeconomic variables. Thus, wage formation is the topic of the last chapter in this text.

## Overview

Chapter 1 on pricing in markets for goods or services notes, firstly, that firms set prices which include mark-ups on direct (variable) costs in order to cover indirect (fixed) costs and some

profits and, secondly, that profit-seeking firms also attempt to exploit an inelastic demand for its products when they can. But this chapter also explains price competition and price leadership in markets where buyers take prices as given and prices are set by firms, as in most consumer markets. Thus, even if not all firms can be price takers, all firms but one can. This model, first suggested by Kenneth Boulding (1941), explains not only why markets sometimes clear but also why they often do not clear or, in other words, why production often is restricted by sales at prices set by firms.

Chapter 1 also clarifies the distinction between consumer markets and other markets, including in particular markets with sealed bidding enforced by a big buyer and markets where market clearing is implemented by “match makers”, as in markets for commodities. It concludes by relating prices to wages and technology in a model which is used repeatedly in the rest of the book, and which shows how prices, production and employment depend on wages, import prices, technology and mark-ups, but also on product demand. Thus, this model also shows that the traditional model of labour demand as a function only of the real wage (and the production function) is incomplete. A complete model of labour demand provides not only a solid microeconomic foundation for a theory of employment – along the lines first suggested by John Maynard Keynes (1936) – but is also easier to understand intuitively, by relating labour demand to all its determinants, including product demand.

Chapter 2 introduces “macroeconomic relations”, that is, monetary restrictions on the aggregate behaviour of economic agents. They show how aggregate profits in the market sector depend on investment, dividends, export surplus, and excess borrowings (borrowings minus savings) in the household sector and the public sector, a relation first suggested by Kalecki (1971). They also imply that net increases of financial assets in the economy sum to zero, a balance condition with many applications. If, for example, the market sector and the household sector together run a financial surplus (savings exceed borrowings), then the public sector must necessarily run a deficit if exports equal imports.

Chapter 2 also shows that the economy’s output in monetary terms can be measured not only by expenditures on final goods and services but also by incomes. It discusses the determinants of the components of output, and argues that the most appropriate point of departure for specifying a “consumption function” – which relates consumption to its most important determinants – is not output interpreted as income but the budget constraint of the household sector, relating spending to incomes, savings and borrowings.

Since inflation means increasing prices, Chapter 3 on inflation begins by summarizing price formation and in particular the determinants of price changes according to Chapter 1.

We note that there are two possible “regimes” for price formation in an industry, and which regime rules depends on the relation between demand and capacity. If firms are producing at full capacity, then higher prices depend on higher demand. If, on the other hand, production is restricted by sales at prices set by firms, then inflation depends on the evolution of wage costs, costs of other variable inputs, mark-ups, and labour productivity, but not on demand.

In Chapter 3 we shall also see how increases in the supply of money relate to inflation. Lending by commercial banks increases the stock of money and may, when borrowings are spent, also increase the demand for consumer goods. But the effect on inflation is substantial only when production is restricted by capacity, since when production is restricted by sales at prices set by firms, demand affects production and employment but not prices.

Chapter 4 on employment shows how a firm’s employment during a year, when the capital stock is given, depends on the demand for its output at the price it sets. While employment is determined by production and labour productivity, and production is determined by capacity or sales at prices set by firms, sales are determined by spending by households or firms. Chapter 4 also includes an example of a formal model of the determinants of employment within the framework developed in Chapters 1-2. Moreover, Chapter 4 shows how the effect of matching problems on employment – and hence also on unemployment – can be measured in vacancy surveys, and how large the effect is.

Chapter 5 on unemployment emphasizes that changes in unemployment, including its level, are caused by changes in employment and changes in the number of people without employment looking for jobs. We note, in particular, that the stock of unemployment is not determined by, but is a *restriction* on, the flow rates. In other words, high unemployment is not caused by long spells of unemployment but vice versa: long durations are caused by high unemployment.

Chapter 6 on wage formation begins by emphasizing that even if real wages – wages adjusted for inflation – certainly matter, we need a theory for how money wages are set in practice. And then not only *real* but also *relative* wages matter, since it is *relative* wages which affect individuals’ choices between occupations, industries, and firms. Moreover, I focus on how wages are *revised*, and in particular on how the *wage level* changes. And to explain the motives for employees’ wage claims and employers’ wage offers, I first discuss two hypothetical and polar cases, one where wages are set by workers or their unions and one where wages are set by firms or their organisations.

Since money is a prerequisite for the division of labour in a society, this concept is fundamental to all parts of economics. It is the basic unit of measurement in this text, as in

Keynes (1936 p. 41). Not even “real” values, adjusted for inflation, can be defined without it, and relative prices are in practice determined by nominal prices set by firms. Moreover, most money is created by commercial banks’ lending, a fact with important consequences for the real economy. For an introduction to money and the financial system, see e.g. Farm (2017d).

This text is written for students and teachers interested in the foundations of macroeconomics as presented in introductory textbooks. But it assumes no prior knowledge of economics, excluding, of course, those sections which briefly relate the text to the literature. This means that it should be accessible also to policy makers and other professionals interested in the workings of a market economy. It also means that the following narrative is not burdened with discussions of current theories or textbooks, even if the facts I rely on are taken from existing literature. This delimitation will make the text not only much shorter but also less abstract and easier to digest.

## Chapter 1: Price formation

A market economy is characterized by voluntary exchange of goods or services for money at prices accepted by both buyers and sellers. But how are prices set? Cost-plus prices are important because they are used at least as a point of departure for pricing in most markets. And the basic reason for them is that a firm sets prices to cover not only direct (variable) costs but also indirect (fixed) costs and some profits, as elaborated in Section 2. Sometimes, however, a profit-maximizing firm tries to exploit an inelastic demand for its products, as discussed in Section 3 on value-based pricing. But we shall also see, in Section 4, how cost-plus prices or value-based prices can be modified by price competition in markets where buyers take prices as given and prices are set by sellers, as in most consumer markets.

Prices are related to wages and output is related to employment in Section 5. The result is a base model of pricing, production and employment which will be used repeatedly in the following chapters for consumer markets. Pricing in other markets are discussed briefly in Section 6, including commodity markets, where prices are set by a market-clearing process, and markets where a big buyer enforces sealed bidding. In the last section my approach to pricing is related to the literature.

### 1.1 Assumptions

There is an important distinction between markets with *production before sales* (production to stock) and markets where *sales precede production* (production to orders). In markets with production to orders firms only produce what they can sell at the prices they set. Services are, of course, almost always produced to order. Otherwise production to orders is possible whenever buyers can accept some waiting time between purchase and delivery. If, for example, buyers want to inspect a car before purchase, they will prefer show rooms where cars are demonstrated, but they may accept some waiting time before a car is produced and delivered from a factory, implying production to orders.

However, in many markets production does precede sales, especially in consumer markets, where customers usually have to visit shops to find what they want to buy. And then a firm has to anticipate its sales at the price it sets. In this case production will in general differ from sales and the difference will change the firm's inventories. But we can often assume that such changes are negligible, so that production equals sales even in markets where production precedes sales, at least approximately.

We also assume that a firm's marginal cost is constant up to a certain level of production – its capacity – where it becomes so strongly increasing that its potential output of goods can be approximated by its capacity even for high prices. A firm's marginal-cost function is then characterized by two parameters: marginal cost and capacity. This is probably not only a useful first approximation but also rather realistic.<sup>1</sup>

Of course, in general a firm's supply curve – that is, its potential output as a function of its price assuming that it can sell all it wants to produce at this price – is not vertical for high prices, even if it usually is rather steep due to constraints on employment in current premises and with current machinery and restrictions on overtime etc. Capacity is consequently in general not a parameter but an increasing function of the price. However, assuming a constant capacity simplifies the analysis considerably without changing its substance.

I also assume that consumers are free to choose between producers (consumer sovereignty), and that firms are free to affect these choices by marketing. And I do not exclude the possibility that a consumer can be affected by other consumers, even if it won't be necessary to model such influences here.

## 1.2 Cost-plus pricing

A firm sets prices to cover costs and obtain some profits. A lower limit is, of course, its marginal cost, since with a price below marginal cost a firm will minimize its losses by reducing output to zero. However, to cover not only variable (direct) costs but also fixed (indirect) costs a firm must set prices above marginal cost, which means that firms in practice set prices as mark-ups on marginal costs.

A basic version of cost-based pricing is cost-plus pricing, which means that firms set prices to cover costs and “normal profits”. More precisely, the cost-plus price  $p^+$  is determined by

$$(1) \quad p^+ = c + mc ,$$

where  $c$  is marginal cost and the mark-up  $m$  is determined so that the firm can cover fixed costs  $f$  and obtain some “normal” profits  $\pi_n$  for its estimated sales  $q$ ,

$$(2) \quad p^+ q - cq - f = \pi_n .$$

It follows that

$$(3) \quad m = (f/cq) + (\pi_n/cq) .$$

---

<sup>1</sup> See, for example, Layard et al. (1991 p. 340), Blinder et al. (1998 p. 102), and Lavoie (2014 pp. 147-56).

Thus, the mark-up is obtained by summing two ratios, namely the ratio of indirect costs to direct costs and the ratio of normal profits to direct costs for estimated sales. But while the first ratio is determined by fixed costs (and estimated sales) the second ratio defines a “normal rate of return on variable capital”. Assuming this to be  $r_n$ , for example 10 per cent, as for many firms examined in a classical article by Hall and Hitch (1939 p. 19), normal profits will be determined by  $r_n$  and direct costs for estimated sales,  $\pi_n = r_n cq$ . Alternatively, “normal profits” are defined as  $\pi_n = rK$ , where  $K$  is the firm’s total capital and  $r$  a “normal rate of return on total capital”; obviously,  $r = r_n (cq/K)$ . Cost-plus pricing also implies that indirect costs and normal profits are allocated to a firm’s products in proportion to their direct costs.

Note that the mark-up  $m$  depends not only on “normal profits” but also on the relation between fixed costs and variable costs. Suppose, for example, that fixed costs increase less than variable costs when output increases between firms in an industry. Then a big firm will have a smaller mark-up than a small firm, other things equal. Note also that direct costs include the costs of intermediate goods and direct labour, while the costs of indirect labour (managers, supervisors, administrators, etc.) are included in indirect costs, together with costs for real capital like premises and machines.

Now, suppose first that the firm expects its sales to be equal to  $q$  independent of the price it sets. Then the firm will also, of course, predict that cost-plus pricing will yield normal profits. (But in this case the firm should also predict that it can obtain even higher profits, by increasing the price further.) Moreover, with actual sales equal to  $s$  the firm’s profits will in general be

$$(4) \quad \pi = (p^+ - c)s - f = mcs - f = (f + \pi_n)(s/q) - f = (s/q - 1)f + (s/q)\pi_n,$$

so that the firm also makes more than normal profits if actual sales exceed estimated sales.

Suppose next that sales are decreasing in price,  $s = D(p)$  with  $D'(p) < 0$ , and that  $p^m$  maximizes profits, as elaborated in Section 3. In this case profits can be at most equal to

$$(5) \quad \pi^m = (p^m - c)D(p^m) - f,$$

which means that normal profits are attainable only if  $\pi_n \leq \pi^m$  and only if  $p^+ = c + mc$  is sufficiently close to  $p^m$ . (Of course, if  $\pi_n = \pi^m$  normal profits can only be obtained for a mark-up  $m$  such that  $c + mc = p^m$ .)

Cost-plus pricing is a common pricing procedure in a market economy.<sup>2</sup> It can sometimes be interpreted as the first stage in a two-stage process. If cost-plus prices yield normal profits the firm knows that there are at least some prices (and sales) which make its profits acceptable. But a rational firm should also realize that its ability to obtain normal profits depends on an assumption which implies that it can generate even more profits – unless it by luck happens to set the profit-maximizing mark-up. Thus, in a second stage the firm may attempt to optimize prices, particularly in a recession with declining sales, when raising prices as suggested by cost-plus pricing may ruin the firm. Even if cost-plus pricing is a convenient rule of thumb in a world of uncertainty, firms may sometimes find it rational to deviate from it.

Cost-plus prices depend on estimated sales, but actual sales will usually differ from estimated sales and production will adjust to actual sales. Prices are not revised daily (as in markets for securities) but kept fixed for some time (e.g. a quarter or even a year), while production adjusts to demand at these prices. And this is true as long as the firm's production is restricted by its sales at the price set by the firm.

But what happens if a firm's production is restricted by its capacity  $k$  and not its sales at its choice of mark-up? To begin with, a queue of customers will develop, making it tempting for the firm to set a price  $p^k$  which clears the market,

$$(6) \quad D(p^k) = k.$$

In practice, however, a firm may hesitate to raise its price if it expects queues to be only temporary, or if it fears that customers facing higher prices than first announced will turn to other firms with excess capacity. Moreover, a firm can avoid problems like these by adjusting its capacity to the variability of its sales. And it seems to be an empirical fact that firms plan some excess capacity in order to avoid losing or antagonizing customers.<sup>3</sup>

Mark-ups and hence cost-plus prices depend on estimated sales ( $q$ ) in the near future, according to (3). The flow of sales normally fluctuates during a year, between months and sometimes strongly between seasons. Sales may also fluctuate between years, with or without a trend. A firm must consequently find out when and how to revise estimated sales in view of the evolution of actual sales and other information.

If the firm believes that the expected value of yearly sales is constant (independent of time), it will think of this as “normal sales” and “normal output”. In this case the firm will

---

<sup>2</sup> See e.g. Scherer (1980 p. 185), Okun (1981 p. 153), Simon (1989 p. 48), Nagle and Hogan (2006 p. 2) and Lavoie (2014 pp. 157-63).

<sup>3</sup> See, for example, Lavoie (2014 pp. 150-54).

plan for “normal output” but will also keep some excess capacity to be able to meet fluctuations of sales above the normal level. However, if the firm not only plans to sell more but also believes that its plans will be realized, it will revise its estimate of normal sales and normal output.

Note finally that if cost-based prices differ so much between firms in an industry that the corresponding price differentials cannot be sustained, it remains to explain why and how prices are adjusted. Thus, cost-plus pricing cannot explain the formation of a *market price* in an industry unless all firms have similar costs.

### 1.3 Value-based pricing

Cost-plus pricing can even be dismissed as a “delusion” in modern management literature, leading to “overpricing in weak markets and underpricing in strong markets”.<sup>4</sup> What is advocated instead is pricing based on “how products and services create value for customers”,<sup>5</sup> or, in other words, profit maximization. It differs from cost-plus pricing by ignoring fixed costs and focusing on the price elasticity of demand. And with profits given by

$$(7) \quad \pi = (p - c)D(p) - f$$

we have

$$(8) \quad d\pi/dp = D(p)(1 - \mu(p)\eta(p)),$$

where  $\mu(p)$  is the contribution margin,

$$(9) \quad \mu(p) = (p - c)/p,$$

and  $\eta(p)$  is the price elasticity of demand,

$$(10) \quad \eta(p) = -pD'(p)/D(p).$$

Thus, a price increase is profitable if the price is “too low”,  $\mu(p) < 1/\eta(p)$ , while a price decrease is profitable if the price is “too high”,  $\mu(p) > 1/\eta(p)$ .

Suppose, for example, that a firm’s customers are insensitive to the price if it is “sufficiently low”, while they rapidly leave the firm for other firms if the price is “sufficiently high”. More precisely we assume that demand is inelastic ( $\eta(p) < 1$ ) for “low prices” and elastic ( $\eta(p) > 1$ ) for “high” prices, and we can even assume that  $\eta(p)$  is increasing in  $p$ .

Then the profit-maximizing price  $p^m$  is uniquely determined by the equation

---

<sup>4</sup> Nagle and Hogan (2006 p. 3).

<sup>5</sup> Nagle and Hogan (2006 p. 27).

$$(11) \quad \mu(p^m) = 1/\eta(p^m).$$

Or suppose that there is a price ceiling  $\hat{p}$  such that demand is inelastic for prices below  $\hat{p}$  but (very) elastic for prices above  $\hat{p}$ , with a discontinuity (“kink”) at  $p = \hat{p}$ . Then the profit-maximizing price is obviously equal to this price ceiling,  $p^m = \hat{p}$ .

Equation (11) is often used to guide modern price management in practice.<sup>6</sup> If, for example, the contribution margin  $\mu$  is 50%, then this margin is also profit maximizing if changing price by 1% is estimated to change sales by 2%. Alternatively, a firm can estimate the profit-maximizing price directly if it finds reasons to believe that demand is inelastic for prices below a certain ceiling  $\hat{p}$  but very elastic for prices above  $\hat{p}$ . And in general, including firms with capacity  $k$ , the profit-maximizing price will be

$$(12) \quad \max(p^m, p^k), \text{ where } D(p^k) = k.$$

Note that while cost-plus pricing has the same mark-up for all products by a firm, value-based pricing implies mark-ups – or more precisely contribution margins – which are inversely proportional to the products’ price elasticities. Moreover, as long as profits are at least normal it may be rational for a monopolist to be satisfied with cost-plus pricing – if further information on demand is too costly. A lower price may appear too risky because the monopolist doesn’t know if demand is sufficiently elastic, and a higher price may also appear too risky because the firm doesn’t know if demand is sufficiently inelastic. After all, estimating product demand, and in particular the price elasticity of product demand, is a difficult problem, particularly for a firm with many products, as in retail.

Note also that pricing according to the formulas above only applies to a monopoly or a small firm assuming that its price revisions will affect only its customers and not its competitors, and that the influence of its competitors’ prices on its sales can be neglected. If value-based prices differ so much between firms in a market that the corresponding price differentials cannot be sustained, it remains to explain why and how prices are adjusted. In other words, not even value-based pricing can always explain the formation of a market price.

#### **1.4 Market price, price competition, and price leadership**

If firms differ so much in an industry that the corresponding differences between cost-plus prices (or value-based prices) cannot be sustained, it remains to explain why and how prices are adjusted. Thus, cost-plus pricing cannot explain the formation of a *market price* in an

---

<sup>6</sup> See e.g. Simon (1989) or Nagel and Hogan (2006).

industry unless all firms have similar costs. Of course, in markets with product differentiation some price differentials may persist, but deviations from a common *price level* cannot be large in equilibrium, after price adjustment (“the law of one price”).

It consequently remains to explain, in a complete theory of pricing, how a market price is established in markets with more than one firm and, in particular, how price competition can arise and affect the adjustment process. And the key to a complete theory of pricing in consumer markets is co-ordination by a price leader.

*Price leadership* means that all firms but one take the price as given or, more precisely, that one of the firms sets a price which the other firms match. And while setting the same price as another firm suggests collusion in markets with sealed bidding, it is both possible and legal in markets where firms are free to observe and revise their prices at any time, as in most consumer markets.

*Collusive price leadership*, maximizing an industry’s profits, presupposes a possibility for firms to jointly fix market prices or market shares. This is usually illegal, but since secret cartels are sometimes exposed we cannot exclude this market form. On the other hand we cannot exclude price competition in markets where a public authority can prevent binding agreements on market prices and market shares.

Now, firms accepting the validity of the law of one price, and operating in a market with a competition authority, will ask what market price they prefer, where a firm’s preferred market price is that market price which maximizes its *individual* profits. If all firms prefer the same market price the choice of price leader among these is immaterial and may be expected to vary randomly or depend on which firm is assumed to have the best information on market conditions. A price leader may in this case be called a *barometric price leader*, following Stigler (1947).

If firms prefer different market prices, due to differences in costs, capacities, or market shares, then the market price will be determined by a *competitive price leader*, that is, a firm preferring the lowest market price, an idea which goes back at least to Boulding (1941). And if there is only one firm preferring the lowest market price, it may be called a *dominant price leader*. Note that such a firm can implement its preferred market price simply by announcing it, while firms preferring a higher market price are forced to follow suit, at least if the price leader has excess capacity. In terms of price adjustment, the market price goes down if *and*

*only if a price cut appears profitable to a firm even if its competitors follow suit. And the market clears if a higher market price would reduce profits for at least one firm.*<sup>7</sup>

In a market with competitive price leadership the problem of a price taker is simple: it sets the same price as the price leader and produces what it can sell at this price or, if its capacity is less than potential sales, what it wants to sell. The problem of a price maker is partly the same as it is for a monopolist, i.e., estimating the industry's product demand and especially its price elasticity. In addition, however, a price leader has to estimate its market share at different market prices, including market prices above the market-clearing level (where demand equals total capacity). And then a firm with a large capacity will sometimes find that not only its market share but also its profits will increase as the market price decreases. A dominant firm may even find it profitable to set a price approaching marginal cost if there are many small firms, as elaborated in Farm (2017a).

In a market where all firms have (approximately) the same cost-plus prices (or value-based prices), not only individual prices but also the market price (price level) will depend on the same factors. And if wages and input prices change for one firm, they will usually change in the same way for all other firms in the industry, and not only individual prices but also the market price (price level) will change in the same way, as if governed by an "invisible hand".

But the "invisible hand" may sometimes be replaced by a "visible hand", particularly during the innovation phase of an industry's product cycle, when an industry often is dominated by a big firm acting as a price leader. In this case the market price is determined by the cost-plus price (or value-based price) of the dominant firm, which, however, may be modified if the dominant firm estimates that decreasing the market price will increase not only its market share (because of the limited capacity of its competitors) but also its profits.

### **1.5 Relating prices to wages and output to employment**

Consider an industry where buyers take prices as given and prices are set by firms, as in most consumer markets – and hence in most markets. In this case price formation depends entirely and directly on demand if production is restricted by capacity in all firms. Higher prices may be associated with higher production, but production can be only marginally higher. On the other hand, if production is restricted by sales at prices set by firms, then prices are determined by firms when they revise their prices according to mark-up pricing,  $p = (1 + m)c$ ,

---

<sup>7</sup> Thus, I don't exclude the possibility of market clearing even if the management literature on cost-plus pricing or value-based pricing implicitly assumes that a firm's production always is restricted by sales (and not capacity) at the price it sets, suggesting that this is a stylized fact for most firms most of the time.

where  $c$  is the direct cost (variable costs per unit of output) and  $m$  is the mark-up on direct costs chosen by a firm.

More precisely, mark-up pricing includes not only cost-plus pricing or value-based pricing as discussed in Sections 2 and 3 but also the mark-up set by a competitive price leader (which is not always the same as the mark-up a monopoly would set) as discussed in Section 4. Note that a price taker in a market with price competition has a mark-up which is determined by the market price  $p$  set by the price leader and the price taker's direct cost  $c$ , so that in this case  $m = (p - c)/c$ . To sum up,  $m$  is a mark-up on direct costs chosen by a firm in order to cover indirect (fixed) costs and some profits, or to maximize profits, or adjust to price competition.

Moreover, direct costs can be related to technology, wages and other direct costs in the following way. Note first that variable (direct) cost  $C$  is a function of both (direct) employment, as measured by  $N$ , and intermediate goods, which we measure by  $M$ , so that

$$(13) \quad C = wN + vM ,$$

where  $w$  denotes the wage level and  $v$  the price level of intermediate goods.

Next we assume that not only employment is proportional to output  $q$ ,  $N = q/a$ , but also other variable inputs,  $M = q/b$ , so that  $vM = vq/b = vaN/b$  and hence

$$(14) \quad C = (w + g)N , \text{ where } g = va/b .$$

In this case even variable costs other than labour costs are proportional to employment, with an addition to the wage level ( $g$ ) which depends on the prices of additional inputs ( $v$ ) as well as the relation between employment and other inputs ( $a/b = M/N$ ). In practice, while  $w$  is measured by labour costs divided by employment,  $g$  is measured by variable costs other than labour costs divided by employment. It follows that direct costs can be written as

$$(15) \quad c = C/q = (w + g)/a = (1 + h)(w/a) , \text{ where}$$

$$(16) \quad g = va/b = vM/N \text{ and } h = vM/wN .$$

We finally assume that not only a firm's capacity ( $k$ ) but also its input technology ( $a/b$ ) is fixed in the short run. Even if a firm in the long run attempts to minimize its direct cost at anticipated input prices by an appropriate substitution between labour and other variable inputs, it can hardly change this mix instantaneously when input prices are revised. A firm can adjust its output prices almost instantaneously to new input prices, but it takes longer to change its technology.

With these assumptions it follows that

$$(17) \quad p = (1 + m)(1 + h)(w/a) \text{ and } N = D(p)/a \text{ if } D(p) < k ,$$

where  $p$  denotes the price the firm sets,  $D(p)$  the firm's sales at this price, and  $m$  is the firm's choice of mark-up.

Note that a firm's production technology is characterized by four parameters, namely capacity ( $k$ ), labour productivity ( $a = q/N$ ), input structure ( $h = vM/wN$ ) and even mark-up ( $m$ ). Labour productivity is a summary measure of the effect of labour on output at current technology and skills, including the use of real capital (machinery) and intermediate goods. Input structure measures the firm's dependence on current production in other firms, while the mark-up on direct costs,  $m = (p - c)/c$ , reflects the firm's dependence on past production of investment goods in other firms. More precisely, if the mark-up only covers fixed costs  $f$ , so that  $m = f/cq$  according to (3), then the mark-up is equal to fixed costs as a proportion of variable costs, and consequently characterizes the relation between 'past inputs' and 'present inputs' by a dimensionless parameter. Note also that employment depends on the demand for the firm's output at the price set by the firm according to (17). This result is quite general according to Farm (2017b), which also shows in detail why the traditional model of labour demand as a function of the real wage is misleading.

The framework in this section will be used as a first approximation or baseline model in the rest of the book. We shall focus on how prices are *revised* when wages, prices of intermediate goods, mark-ups and technology change. In general an industry's output can be used by industries producing its inputs. And then the prices of an industry's inputs depend on the price of its output, so that input prices and output prices can be adjusted several times. The importance of this adjustment problem in an industry can be measured by variable costs other than labour costs measured as a proportion of labour costs ( $h = vM/wN$ ), provided, of course, that the industry's output is used by industries producing its inputs. However, outputs of final goods are not intermediate goods, and we shall focus on the pricing of final goods, in particular consumer goods.

## 1.6 Pricing in special markets

To appreciate the fundamental difference between consumer markets and other markets, I will here briefly discuss *classical markets*, where market clearing is established by an auctioneer; *commodity markets*, where a market-clearing process is organized by 'match makers' or 'market makers'; and *sealed bidding*, where firms have to predict competitors' price offers. I conclude with some brief comments on business-to-business markets.

### *Classical markets*

In markets where production precedes sales producers have to 'bring their goods to a market place' before sales are completed. In modern times this 'market place' is one or several shops in a city or country, or an address on the Internet, while in ancient times, before the industrial revolution, the 'market-place' was a place where all producers of a good had to bring their output in order to sell it. Suppose, for simplicity, that this happened once a month during a single day, and that goods brought to the market-place must be sold during this day, for example because the products brought to the market-place were perishable. How will such a market be organized?

Of course, sellers want prices to be as high as possible while buyers want them to be as low as possible. Now, if producers start by asking high prices they may run the risk that buyers postpone their purchases, hoping that prices will be lower towards the end of the day, particularly since they know that the producers plan to sell everything they have brought to the market-place. Or buyers may start bargaining with sellers, leading to a wide distribution of prices during the day's transactions.

In general prices are consequently indeterminate in a completely unregulated market. Suppose, however, that the sellers attempt to agree on a common market price  $p$  which they announce at the beginning of the market day and stick to until the end of the day. Even buyers may find this convenient, since then they can go to the market at a time which suits them and buy what they want at a price they know beforehand without time-consuming bargaining. But they cannot do this if total supply  $Q$  falls below demand  $D(p)$  during the day so that customers arriving towards the end of the day get nothing.

Thus, both sellers and buyers should be able to find it rational and convenient to let an auctioneer announce a market price at the beginning of the market day, that is, a price which applies to all transactions during the day. But both parties will be satisfied only if the market clears, so that  $p = p^o$ , where  $D(p^o) = Q$ .

Of course, if  $p^o > p^o$ , where  $p^o$  maximizes aggregate revenues  $pD(p)$ , then every producer could earn more at a higher market price. But every producer could not be sure of this unless individual sales are proportional to individual supply not only at a market price which clears the market but also at a higher market price. And proportional rationing would not occur spontaneously but presupposes a system of rationing which would inconvenience the customers. Moreover, destroying excess supply of goods at the end of the market-day

would be offensive to most people. Thus, market clearing is a market form which can be supported by both sellers and buyers not only if  $p^o \leq p^e$  but also if  $p^o > p^e$ .

Of course, in practice it would not be possible for an auctioneer to predict total sales so accurately that exact market clearing would follow. Since excess demand would be frustrating to both sellers and buyers, a plausible strategy for an auctioneer might be to aim at some excess supply, but also to add a short period of sales at lower prices at the end of the market period, to get rid of remaining stocks.

Now, if producers know that their products will be sold in the market, what will they produce? Because of market clearing, every producer is able to sell all it supplies to the market, so it is tempting for every producer to increase the supply the next market period. In an expanding economy with increasing demand and restrictions on how fast production can grow, this may also be consistent with a stable market price.

However, with stable demand and ability to expand production fast, rational producers will realize that expanding output also has a negative effect on the market price. In fact, they are caught in a dilemma where every producer tries to predict what the competitors will do – in order to choose a profit-maximizing output given these predictions.

Suppose now that there are  $n$  firms in the industry and that the demand function is linear,  $D(p) = (a - p)/b$ . In this case the market price  $p$  is market clearing when total output is  $Q$  if  $(a - p)/b = Q$  so that the market-clearing price is  $a - bQ$ . Assuming in addition that every firm has the same marginal cost ( $c$ ), the profits of firm  $i$  with output  $q_i$  and fixed costs  $f_i$  is

$$(18) \quad \pi_i = (a - bQ)q_i - cq_i - f_i, \text{ where } Q = \sum_{j=1}^n q_j.$$

Of course, a firm doesn't know beforehand what its competitors' outputs will be. But it can derive *consistent predictions* of all outputs by solving the system of equations

$$(19) \quad \frac{\partial \pi_i}{\partial q_i} = (a - bQ) + (-b)q_i - c = 0, \quad i = 1, 2, \dots, n,$$

since by solving these equations for all outputs every firm is supposed to maximize its profits in the usual way, taking other outputs, or more precisely the predictions of other outputs as given. Moreover, if firms choose outputs in this way, then they will not regret their choices when outputs can be observed, since a firm's output will indeed maximize its individual profits, taking all other outputs as given. The firms' choices can consequently also be characterized as a *self-enforcing agreement*. Note, however, that such an agreement presupposes *perfect information* on market conditions or, more precisely, that all firms know

how many firms there are in the market; that all of them have the same marginal cost; and that they have the same (true) information on the demand function.

Now, to solve (19) we note first that (19) implies that  $q_i$  is constant (independent of  $i$ ) and hence that  $q_i = Q/n$ . Substituting this in (2) we find that

$$(20) \quad Q = \frac{a-c}{b} \frac{n}{n+1} = D(c) \frac{n}{n+1} \quad \text{and} \quad q_i = \frac{D(c)}{n+1}, \quad i = 1, 2, \dots, n.$$

It follows that total output tends towards  $D(c)$  and that the market price consequently tends towards marginal cost  $c$  as the number of firms grows. Note that the competitive effect of an increasing number of firms applies even when the firms are identical, in contrast to what applies to markets where prices are set by firms.

Of course, producers have to ‘bring their goods to a market place’ even in modern markets where production precedes sales. But in modern markets the ‘market place’ is not the same as a classical market-place. And in modern consumer markets firms are *not* caught in a dilemma where firms choose outputs and sales are adjusted to these outputs at a price set by a market-clearing auctioneer. Instead outputs are adjusted to sales at prices set by firms.

### *Commodity markets*

Markets for commodities like agricultural products, oil, electricity, metals and other raw materials are organized through exchanges establishing equality between supply and demand during a day, at least approximately. In practice the market-clearing process is implemented by large international banks acting either as match makers or market makers. In the first case dealers match orders to buy and orders to sell (for a fee). In the second case market makers with inventories of commodities announce a bid price, at which they are prepared to buy commodities, and a somewhat higher offer price, at which they are prepared to sell commodities. Unless sales approximately equal purchases during a day, market makers will either run out of inventories or run out of money, so they have to be good at anticipating demand and supply, or at least quick to adjust their prices to changes in sales or purchases.

However, a supplier of a commodity to an exchange cannot normally – because of imperfect information on competitors and demand functions – predict other suppliers’ quantities during a day. And because of this uncertainty commodity prices are usually volatile, even if variations are moderated by statistical laws in large markets. Uncertainty about future prices may also be reduced by *futures markets*. And to distinguish between

futures markets and the underlying markets for immediate delivery of commodities, the latter are usually called *spot markets*.

Note also that even if a commodity is traded in both a spot market and a futures market, it can also be traded in other forms, for example through special agreements between a big producer and a big user. The market for a commodity may even include price-taking consumers, as modern markets for electricity demonstrate, suggesting the possibility of competitive price leadership even in commodity markets.

### *Sealed bidding*

Sealed bidding means that sellers have to submit price offers to a buyer for an undertaking specified by the buyer committed to a contract with the cheapest bidder. The buyer may be a big firm or a government agency. The undertaking may be the construction of a building or the operation of a railway or subway or, in general, the realization of some project. The purpose with sealed bidding is, of course, to attract bids from the most efficient producers but also to prevent such producers from asking excessive mark-ups.

In the long run competition enforced by sealed bidding may also increase an industry's efficiency. However, if one of the firms is so efficient that it wins all bidding contests, then its competitors may leave the industry and the industry becomes a monopoly. On the other hand, if there are several firms in an industry which are equally efficient, and the firms know this, sealed bidding implies a dilemma for the firms.

To see this, suppose that every firm estimates the cost of a project to be  $c$ , including fixed costs. Suppose also that every firm realizes that the buyer is prepared to pay more, say  $c + d$ . Now, a firm which predicts that its competitors will bid  $c + d$  will bid somewhat less in order to get the contract. But then the firm realizes that its competitors may reason similarly, so it reduces its bid even further – and further. Or, having taken a course in game theory, the firm's CEO realizes that the only rational predictions imply that every firm bids  $c$ , with a probability equal to  $1/n$  of getting the contract if there are  $n$  firms in the market.

While this is nice for the buyer, the sellers prefer obtaining  $c + d$  with probability  $1/n$ . But they cannot obtain this outcome if all of them bid  $c + d$ , since setting the same price as another firm smacks of collusion in markets with sealed bidding. Instead they can attempt to secretly agree on taking turns in being the lowest bidder at  $c + d$ , provided that there are many opportunities for bidding in the industry, so that all firms can share the profits from

collusion in the long run. On the other hand, such agreements cannot be enforced legally and are therefore unstable, particularly if a very large contract is offered.

Another problematic aspect of sealed bidding is that a firm which have obtained a contract can increase its profits by cutting costs below the costs assumed when the bid was submitted. This possibility may of course stimulate the development of more efficient technology or organisation, but it may also tempt the firm to reduce quality or wages by unorthodox methods and hence make it necessary to introduce costly surveillance.

Thus, sealed bidding is a market form with many complications. It may appear to benefit the buyer in all circumstances, which may be a reason for making it mandatory for government agencies when buying goods or services or outsourcing activities. But it can also have drawbacks for the buyer, and its competitive consequences can be neutralized by the sellers. It is certainly a market form which deserves attention in economics. But it is not applicable to consumer markets.

### *Business-to-business markets*

Excluding sealed bidding enforced by a big buyer, how are prices set in trade between firms? There may be important differences between markets for intermediate goods and markets for investment goods. We may have price leadership even in some business-to-business markets, especially for standardized goods produced by large firms. On the other hand, subcontractors may have to accept prices set by buyers. Even if firms announce list prices for their products, they may regularly enter into bargaining with firms offering large orders. In such negotiations we can assume not only that cost-plus prices define a floor for the seller but also that the seller withholds such information from the buyer, while the seller's problem is to guess what the buyer is prepared to pay. Perhaps this is all that can be ascertained by outsiders about prices set through bargaining.

## **1.7 Relation to the literature**

In this book I follow a *behavioural approach* to price formation, focusing on actual pricing procedures, instead of an *equilibrium approach*, focusing on how prices are determined “in equilibrium” by a set of equations. And I do this not least because an equilibrium approach is related to a completely different question.

A fundamental problem in economics is undoubtedly if there exist prices such that all consumers can buy what they like at these prices and at the same time all producers can sell what they like. Walras suggested that the answer is “yes” and also that all outputs and prices –

but only *relative* prices – are determined simultaneously in such a “perfect economy”, as the solution to a system of equations. It was eventually realized, however, that even the *existence* of such a “perfect economy” required some rather strong assumptions, as documented, for example, in Arrow and Hahn (1971). In any case, this has nothing to do with how prices are set in the real world – unless one can show how actual prices are adjusted to equilibrium prices without a global auctioneer.

Now, even if a “perfect economy” cannot be realized, the following modification does exist. Suppose that prices are set by producers, and that all consumers can buy what they want at these prices. Thus, while firms determine prices, consumers determine quantities, as in most consumer markets. This is not a “perfect” economy, since firms do not necessarily produce all they want to produce at the ruling prices. On the other hand, when firms restrict production to sales, they do this at prices they have chosen voluntarily. Moreover, during trade both buyers and sellers take prices as given, implying that all of them accept the outcome of the pricing process.

To cover not only variable (direct) costs but also fixed (indirect) costs and some profits, a firm must set prices above marginal cost, which means that firms in practice set prices as mark-ups on marginal (direct) costs. The mark-up may be chosen to cover fixed cost and normal profits for estimated sales (cost-plus pricing), or to maximize profits with respect to a perceived individual demand curve (value-based pricing). If, however, firms have different cost-plus prices – or value-based prices – the market price will be determined by the cost-plus price (or value-based price) of a price leader, which, however, may be modified if the price leader estimates that decreasing the market price will increase not only its market share (because of the limited capacity of its competitors) but also its profits. And in a market with price leadership a price taker has a mark-up  $m$  which is determined by the market price  $p$  set by the price leader and the price taker’s direct cost  $c$ , so that in this case  $m = (p - c)/c$ .

Note that in practice *relative* prices on consumer goods are determined by the relation between *nominal* prices set by firms, conditional on wages and prices on imports. But even if prices are based on costs, they are also affected by demand. In other words, relative prices depend not only on relative costs but also on relative utility of goods. However, the relative utility of goods, as judged not by a single consumer but by all consumers, is mainly reflected by relative sales at prices reflecting the relative costs of producing them.

Note also that *monopolistic competition* is not a plausible model of pricing in practice. While perfect competition presupposes that *all* firms are price takers, monopolistic

competition presupposes that *no* firm is a price taker. More precisely, prices are set independently by all firms assuming that every firm profit-maximizes relative to an individual product demand function contingent on the other firms' prices. However, since prices are set independently, they do not necessarily result in the same price or – in markets with product differentiation – the same price level. Price differentials may lead to price adjustment and convergence to a market price, but then a plausible theory of price adjustment is required.

Using game theory, instantaneous adjustment is theoretically possible, but a Nash equilibrium in pure strategies does not always exist (not even in markets with product differentiation). And when it does exist, its derivation presupposes perfect information on demand functions and cost functions (see Benassy 1991 and Appendix 2 in Farm 2017a). Note also that while monopolistic competition is based on *objective* demand functions, value-based pricing is based on *subjective* demand functions and can consequently be characterized as “monopolistic competition in practice”. But note that value-based pricing is only applicable to some products, as discussed above.

### *Price stickiness*

“Price stickiness” is often thought to be a necessary prerequisite for Keynesian economics. It is also obvious that cost-plus pricing, supplemented by price leadership, predicts sticky market prices, since wages, prices of other inputs, labour productivity, and mark-ups do not change that often. The stability of prices in consumer markets and industrial markets are by now also well documented, for example in Blinder et al. (1998) and Fabiani et al. (2007).

“Price stickiness” in a market implies that prices move more slowly than prices in financial markets or commodity markets, suggesting that prices are not market-clearing. But economic theory should reflect the fact that consumer markets and industrial markets are *not* organized in the same way as financial markets or commodity markets. Moreover, if production is restricted by sales, then capacity has no influence on market prices, while increasing demand will increase production and employment but not market prices. Thus, the basic question in economics is not why prices, once chosen, can be “sticky”, but why firms choose prices such that production is restricted by sales. And the answer offered here is simply that it often – perhaps most of the time – is profitable for firms to do so.

If demand in an industry is so large that production is restricted by capacity at cost-based prices, then variation in demand can only affect prices (or delivery lags). But otherwise, when production is restricted by sales at prices set by firms, prices will only adjust to variation in costs (and labour productivity), not sales. The optimal timing of a price revision may in theory

be an intricate problem, particularly when a change in costs is expected to be small or temporary. In practice, however, prices are often only revised when wages are revised or prices of important raw materials (like oil) are changed.

*On benchmark models of pricing*

It is often argued that effects of the properties of a market are best characterized as deviations from a well-understood benchmark.<sup>8</sup> I agree, but I choose perfect competition – or more precisely market-clearing – as a benchmark only for financial markets and commodity markets, while I choose competitive price leadership (CPL) as a benchmark for consumer markets and industrial markets.

Thus, even if a financial market or a commodity market in practice is organized by big banks as a matching process (matching offers to buy with offers to sell), a model with a single auctioneer matching supply to demand is a useful benchmark. And in markets where consumers take prices as given and prices are set by producers, coordination by a competitive price leader is a useful benchmark.

Thus, in markets with differentiated goods, price differentials can fruitfully be discussed as deviations from the price level predicted by CPL. And in markets where firms independently set approximately the same price (due to similar costs), it can be argued that the common price level is determined “as if” announced by a (barometric) price leader. Moreover, in markets with sealed bidding and in other business-to-business markets, deviation from CPL may be an informative characteristic.

---

<sup>8</sup> See e.g. Blanchard and Fischer (1989 p. 27).

## Chapter 2: Macroeconomic relations

This chapter explores the origin of profits: how profits as residual income are generated by a single firm and in the aggregate by the market sector in relation to the household sector, the public sector and the foreign sector. It also introduces *macroeconomic relations*, that is, monetary restrictions on the aggregate behaviour of economic agents. Some of the restrictions are obvious accounting identities, but some are less transparent, even if they are derived from accounting identities. Since these derived relations have interesting interpretations and important applications I prefer to call them “relations” instead of “identities”.

Section 1 defines a firm’s profit as residual income and relates a firm’s investment to its retained profits, borrowings and savings. Section 2 begins the transfer from microeconomics to macroeconomics by first aggregating over all firms in the market sector and then relating the market sector to the household sector, the public sector and the foreign sector. Section 3 introduces a fundamental relation between net savings in the four sectors of an economy, while Section 4 shows how aggregate profits in the market sector depend on aggregate investment in the market sector and excess borrowing (borrowing minus saving) in the three other sectors. Gross domestic product (GDP) is defined in Section 5, while the determinants of consumption and investment are discussed in Sections 6 and 7. Section 8 adds some brief comments on economic growth, that is, growth of total output as measured by GDP at constant prices. The last section relates the text to the literature.

### 2.1 Profit and investment

We start from a firm’s *residual income* ( $ri$ ) during a year (or quarter) defined as

$$(1) \quad ri = r - wn - vm - f ,$$

where  $r$  denotes revenues from sales of goods or services,  $n$  employment,  $w$  the wage level,  $m$  intermediate goods used in production,  $v$  the price level of intermediate goods, and  $f$  fixed costs, including interest paid for debts, rent for premises, and leasing costs for machinery, but also costs for mandatory maintenance of premises and machinery.

The classification of maintenance costs is notoriously difficult, because maintenance can sometimes be postponed without hurting current production, while maintenance sometimes may imply improvement of real capital. In any case maintenance costs are not part of variable costs (proportional to output), so it can always be classified as either fixed costs (if it cannot be postponed) or investment.

It won't be necessary to specify price setting in this chapter. However, since revenues are obtained by sales at prices accepted by buyers, and costs are payments to sellers at prices accepted by sellers, residual income can be interpreted as a measure of the value to society of a firm's transformation of resources into goods or services. Residual income is also an important incentive for a firm's owners and managers, since it is only residual income which after taxes and retained profits can make them seriously rich through dividends or bonuses.

Profits are defined as

$$(2) \quad \pi = ri - t,$$

where  $t$  denotes taxes. Taxes are in practice functions of sales or (taxable) profits or wages or value added, but we don't have to specify the details at this stage. But note that taxes include all taxes collected from the firm, not only tax on (taxable) profits. Moreover, profits are distributed as dividends ( $d_d$ ), bonuses ( $d_b$ ) and retained profits equal to

$$(3) \quad \pi - d = r - wn - vm - f - t - d,$$

where  $d = d_d + d_b$  includes not only dividends to owners but also bonuses to managers.<sup>9</sup>

Now, a firm's activities include not only production of goods or services which generates revenues and current profits when sold, but also purchases of real capital in order to sustain or increase future profits. Thus, *investment* includes purchases of premises and machinery as well as other goods and services which are used not as input to current production but to sustain or increase capacity or reduce variable costs (increase productivity) in the future or develop new products through research and development. Note that investment includes replacement of worn out or outdated real capital (in order to sustain future profits) and also payments for additional maintenance of existing premises and machinery, apart from mandatory maintenance costs (which are classified as fixed costs).

Investment in real capital during a year is financed by retained profits during the year, new debt or dissaving. Retained profits can alternatively be saved in order to buy real capital in the future. Saving may be followed by buying securities and dissaving may be preceded by selling securities. The possibility to finance investment by dissaving depends, of course, on the stock of bank deposits or securities in the beginning of the market period, which depends on past profits, past dividends and bonuses, past borrowing, past investment, and ultimately on bank deposits supplied by buyers of the firm's stock issues.

---

<sup>9</sup> Note that here bonuses only include remunerations which are based on profits. Other kinds of performance pay are included in the wage bill.

Thus, investment ( $i$ ) and saving ( $s$ ) by a firm during a year are financed by its retained profits ( $\pi - d$ ) during the year and borrowing ( $b$ ) according to

$$(4) \quad i + s = \pi - d + b,$$

which we can interpret as a budget constraint for a firm's investment in real capital.

Repayment of debt is registered as negative borrowing.

Equation (4) shows how money generated by current profits and new debt can be used for investment in real capital ( $i$ ) and savings ( $s$ ), including investment in securities. For a firm without savings and borrowings ( $s = b = 0$ ) we have  $i = \pi - d$ , which shows how all the cash generated by retained profits during a year is used to buy investment goods. For a firm without current retained profits and savings ( $\pi - d = s = 0$ ) we have  $i = b$ , indicating that all of the firm's investment is financed by borrowing. And for a firm with dissaving (negative  $s$ ) we have  $i = \pi - d + b - s$ , which shows how investment in general is financed by current retained profits, borrowing, and dissaving, where dissaving depends on past retained profits.

Note that savings include investment in securities or, more precisely, savings may be followed by purchasing securities – and dissaving may be preceded by selling securities – even in production firms. The firm may even be trading securities, with profits or losses from such activities. However, even if such trading can have important consequences for the distribution of wealth and liquidity between firms I will not attempt to model this here.

Real capital has a cost, but the use of real capital can affect the profits and the balance sheet of a firm in different ways. First, the use of real capital can take the form of renting premises or leasing machines. This will affect the firm's profits directly, hopefully by increasing revenues more than it increases costs. But it will not affect the firm's balance sheet, nor will it be registered as investment.

Otherwise the use of real capital takes the form of purchases of investment goods, for example new machines replacing old machines. If the purchase is financed by borrowing, profits as defined here (in terms of cash flows) will be reduced by the interest on the loan until the loan is repaid. But profits will also be reduced by repayment of the loan, either by the entire loan after some years (without amortization) or piecemeal every year until final repayment (with amortization). On the other hand, if the investment is financed by dissaving based on past retained profits, profits as defined here will be reduced by the entire investment the same year the investment is made, while higher profits the following years will reflect the return on this investment.

How will the investment affect the firm's balance sheet? Normal practice is to reduce the value of the investment each year by "depreciation", which also is treated as a cost which

reduces a firm's "taxable profits". This also applies when investment is financed without borrowing. However, "depreciation" will affect profits as defined here only by affecting taxes. Of course, the purchase of real capital will always reduce a firm's *liquidity*, either at once, if payment is made without the support of new borrowing or otherwise when the loan is repaid.

## 2.2 Aggregation and relations between sectors

In this section we begin the transfer from microeconomics to macroeconomics by first aggregating over all firms in the market sector and then relating the market sector to the household sector, the public sector, and the foreign sector.

### *The market sector*

Summing (1) with  $ri = \pi + t$  over all firms we obtain

$$(5) \quad \Pi = R - W_m - Z - F - T_m ,$$

where  $\Pi$  denotes aggregate profits,  $T_m$  aggregate taxes and  $W_m$  aggregate wages in the market sector,  $W_m = w_m N_m$ , where  $N_m$  denotes aggregate employment and  $w_m$  the wage level in the market sector. But summing (1) over all firms also means that we eliminate cash flows between firms which are revenues for one firm (and consequently part of  $r$ ) and costs for another firm (and consequently part of  $vm$  or  $f$ ). After aggregation payments for intermediate goods ( $Z$ ) consequently denote payments to the foreign sector for imports,  $Z = vM$ , where  $M$  denotes the volume of imports and  $v$  its price level. And aggregate revenues ( $R$ ) include spending on goods or services by the household sector, the public sector and the foreign sector but also the spending on investment goods (real capital) by domestic firms.

Note, in particular, that a firm's purchases of real capital from another firm are revenues (part of  $r$ ) for the latter firm, but are not registered as costs of intermediate goods (part of  $vm$ ) by the former firm. Note also that most of firms' fixed costs – including interest payments for debt, rents for premises and leasing costs for machinery – are also revenues for other firms, including banks, so that  $F$  only includes interest payments to other sectors. For simplicity we assume here that only the household sector receives such payments, for example interest on corporate bonds.

Revenues from selling final goods or services are determined by *effective demand*, that is, spending on final goods and services produced by the market sector. And such spending includes expenditures by households on consumption goods ( $C$ ), expenditures by public

institutions on goods from the market sector ( $G$ ) and expenditures by foreigners on export goods ( $X$ ), but also expenditures by firms on investment goods ( $I$ ), so that

$$(6) \quad R = E \text{ , where}$$

$$(7) \quad E = C + I + G + X \text{ ,}$$

where  $E$  denotes total expenditures on final goods or services produced by the market sector (aggregate effective demand). Note that we here focus on expenditures on output from the market sector, while expenditures by the public sector on wages paid to households will be introduced below.

Effective demand is not the same as demand and the distinction is important. Demand refers to the quantity of a particular good purchased at a particular price. The concept is related to a buyer's problem of choosing between different goods at given prices. *Effective* demand, on the other hand, is related to an agent's *budget*, i.e. its plan to allocate its available income and money holdings to various broad purposes or categories, like food, clothes, housing, leisure and saving if the agent is a household. Thus, while demand is measured in physical units, presupposing explicitly or implicitly a certain price for a certain good, effective demand is measured in monetary units.

The market sector contributes to expenditures on final goods through investment. Investment depends, of course, on the supply of capital goods, or more precisely on the ability of some firms to produce investment goods instead of consumer goods. Capacity restricts production and determines at what level of production increasing effective demand raises prices but not production. And unless capacity restricts production of investment goods, demand does.

Moreover, summing (4) over all firms we find that aggregate investment ( $I$ ) and aggregate savings ( $S_m$ ) in the market sector are financed by retained profits ( $\Pi - D$ ) and borrowings ( $B_m$ ) according to

$$(8) \quad I + S_m = \Pi - D + B_m \text{ ,}$$

where  $D$  denotes aggregate dividends and bonuses. Some bonuses may be paid to foreigners, but for simplicity we assume here that all of them are paid to domestic households. We also assume that all dividends are paid to domestic agents and (for simplicity) only to households. Since borrowings can be negative (implying repayment of debt), summing over all firms means that  $B_m$  equals net borrowings by firms, that is, new debt less repayment of debt.

### *The household sector*

Households obtain wages from employment in the market sector ( $W_m$ ) and the public sector ( $W_p$ ), interest payments ( $F$ ) and dividends and bonuses ( $D$ ) from the market sector, and financial assistance ( $A$ ) from the public sector (benefits from the social insurance system minus interest on government bonds owned by households). Households also obtain money from borrowings ( $B_h$ ). I define *consumption* as expenditures by households on final goods produced by the market sector, including purchases of new houses or new apartments.

Consumption ( $C$ ), taxes paid by households ( $T_h$ ) and savings made by households ( $S_h$ ) are financed by labour income, dividends and bonuses, financial assistance and borrowings according to the budget constraint for the household sector,

$$(9) \quad C + T_h + S_h = W_m + W_p + F + D + A + B_h.$$

Note that consumption includes payment for all services bought from the market sector, even interest paid for mortgages and other financial contracts, while  $F$  includes interest paid by firms to households owning corporate bonds. This equation can also be obtained by summing budget constraints over all households or individuals, even if the budget constraint for a single individual as a rule has fewer terms. (An individual considering dissaving is, of course, also restricted by available bank deposits and holdings of securities.)

### *The public sector*

The public sector consists of non-profit organizations, some performing state functions (parliament, government, courts, defence, police, central bank, etc), some administering taxes and financial assistance (unemployment insurance, pension, social security, etc), and some producing services (for example health care or education).

Services in the public sector are not sold but distributed “on demand”, possibly for some fees which we include in taxes. State functions and public services are produced by labour from households (employment  $N_p$  with wages  $W_p = w_p N_p$ ) using inputs from the market sector (government spending  $G$ ). Expenditures on public goods and services ( $W_p + G$ ), financial assistance ( $A$ ), and savings ( $S_p$ ) are financed by taxes from the market sector ( $T_m$ ), taxes from the household sector ( $T_h$ ), and borrowings ( $B_p$ ), so that

$$(10) \quad G + W_p + A + S_p = T_m + T_h + B_p.$$

Taxes depend on output and employment in the whole economy, while financial assistance depends on unemployment, disability and old age, etc. Employment and services in the public sector may be determined in real terms by political decisions, or they may depend on fiscal rules restricting expenditures, taxes, and borrowings. Note that expenditures on public goods can be “investment” in the sense that they provide utility over many years (like expenditures on railways, roads, schools and other infrastructure). Since borrowings can be negative (meaning repayment of debt), summing over all public institutions means that  $B_p$  equals net borrowings by the public sector, while  $B_p - S_p$  is the consolidated budget deficit of the public sector.

### *The foreign sector*

Foreign purchases of domestic goods or services ( $X$ ) or domestic assets ( $S_f$ ) generate foreign exchange which is used to buy imports ( $Z$ ) or foreign assets ( $B_f$ ) or increase foreign exchange reserves ( $\Delta FER$ ), so that

$$(11) \quad X + S_f = Z + B_f + \Delta FER.$$

Note that we can interpret  $S_f$  as savings by foreigners and  $B_f$  as borrowings by foreigners.

Even if all domestic agents (banks, firms, and households) can have holdings of foreign exchange, most of them are kept by the central bank, so that without intervention by the central bank in the forex market we have (at least approximately)  $\Delta FER = 0$  and

$$(12) \quad X - Z = B_f - S_f.$$

Thus, a current account deficit ( $X - Z < 0$ ) must be balanced by a capital account surplus ( $S_f - B_f > 0$ ). In some cases eq. (12) may show unequivocally how a current account deficit is financed. As an extreme example, suppose that, at a given exchange rate, all exports are financed by lending money to foreigners and all imports are financed by borrowing money from foreigners. In this case a current account deficit is directly and automatically financed by a capital account surplus.

Suppose, on the other hand, that a country’s government bonds become so attractive to foreign investors that it implies an excess demand for the domestic currency at the current exchange rate. This will raise the value of the domestic currency in terms of the foreign currency and consequently make imports cheaper to domestic residents and exports more expensive to foreigners. If the appreciation of the domestic currency does not stop until the net outflow of payments for goods (a current account deficit) equals the net inflow of

payments for domestic assets (a capital account surplus), then it may appear as if a current account deficit is financed by a capital account surplus when in fact the current account deficit is caused by a capital account surplus.

In a world with free capital flows between countries, a small increase in domestic interest rates above foreign rates may also sometimes be sufficient to attract capital and create a capital account surplus which can finance a current account deficit, at least in the short run. But it cannot attract capital “in equilibrium” by higher interest rates, since capital flows will stop after some time, either if interest rates are equalized or if investors run out of “movable capital”.

Moreover, a “strong dollar” does not necessarily depend on a high interest rate in the US attracting capital from abroad to the US bond market. It may also depend on very high demand for safe and liquid US assets at given interest rates.<sup>10</sup> Note also that attracting capital may not only finance a current account deficit but also increase it by appreciating the currency and hence reduce exports and raise imports. On the other hand, a country can stimulate the outflow of capital by reducing its policy rate. This will depreciate its currency so that exports increase and imports decrease.

### 2.3 The fundamental identity

In the rest of this chapter I will derive and discuss some important consequences of the relations above. First, it is easy to see that

$$(13) \quad (S_m - B_m) + (S_h - B_h) + (S_p - B_p) + (S_f - B_f) = 0,$$

where  $(S_m - B_m)$  is the increase of financial assets in the market sector during a year,  $(S_h - B_h)$  the increase of financial assets owned by households,  $(S_p - B_p)$  the increase of financial assets in the public sector (corresponding to a budget surplus), and  $(S_f - B_f)$  the increase of financial assets owned by foreigners. Intuitively, the sum of these balances must necessarily be 0, since the financial assets of one agent cannot increase without a corresponding decrease in the financial assets of another agent, or, as emphasized by Wray (2012 p. 1), “for every financial asset there is an equal and offsetting financial liability”. A formal proof is given below.

Equation (13), often called the “fundamental identity”, has many applications, as emphasized, for example, by Lavoie (2014 p. 259). For example, a surplus in the public sector

---

<sup>10</sup> Blanchard (2009 p. 449).

( $S_p > B_p$ ) requires a deficit for at least one of the other sectors, for example  $S_h < B_h$  or  $S_m < B_m$  if foreign trade is balanced ( $X = Z$ ). Or large government deficits ( $S_p - B_p < 0$ ) and large current account surpluses ( $X - Z > 0$  and hence  $S_f - B_f < 0$ ) presuppose large savings by the private sector ( $S_m - B_m + S_h - B_h > 0$ ). And if net borrowings by firms are equal to net savings by households, so that  $B_m - S_m = S_h - B_h$ , then we have not only  $X - Z = B_f - S_f$  but also, from (13),  $B_f - S_f = S_p - B_p$ . Thus, a current account deficit can sometimes be equal to the government's budget deficit. And anyone who proposes to cut government deficits must admit that this cannot be done unless the private sector surplus or the foreign surplus is reduced, as emphasized, for example, by Wray (2012 p. 15).

The formal proof is as follows. According to Section 2 we have the following relations:

$$\begin{aligned} I + S_m &= R - W_m - Z - F - T_m - D + B_m, \\ C + T_h + S_h &= W_m + W_p + F + D + A + B_h, \\ G + W_p + A + S_p &= T_m + T_h + B_p. \end{aligned}$$

Summing these relations and substituting  $R = E = C + I + G + X$  according to (6) and (7) we obtain, since several terms cancel,

$$S_m + S_h + S_p = X - Z + B_m + B_h + B_p,$$

and substituting  $X - Z = B_f - S_f$  according to (12) we obtain (13).

Note that (13) is an identity which says nothing about causation. Thus, it remains to see, in a particular application, which term(s) that are exogenously determined, and which term(s) that are (endogenously) determined by (13).

## 2.4 The profit equation

Second, the determinants of aggregate profits are given by

$$(14) \quad \Pi = I + D + X - Z + (B_p - S_p) + (B_h - S_h).$$

This equation follows immediately from the definition of profits as residual income – revenues minus costs – in the production of goods and services in the market sector:

$$\Pi = (C + I + G + X) - (W_m - Z - F - T_m) \text{ according to (5), (6), and (7),}$$

$$\text{with } C = W_m + W_p - T_h + F + D + A + B_h - S_h \text{ according to (9),}$$

$$\text{and } G = T_m + T_h - W_p - A + B_p - S_p \text{ according to (10).}$$

Thus, (14) is not only an identity, but an identity with a causal interpretation, that is, an “equation”. It is also easy to see intuitively why the excess of exports over imports ( $X - Z$ ) contribute to aggregate profits, namely by increasing the revenues of the market sector more than its costs. This also applies to purchases of market goods by the household sector or the public sector which are financed by excess borrowings ( $B_h - S_h$  or  $B_p - S_p$ ).

Moreover, dividends and bonuses ( $D$ ) can add to aggregate profits since they can add to spending on consumer goods without increasing costs in the market sector. However, this potential doesn’t materialize if dividends and bonuses are used to increase savings  $S_h$  or decrease borrowings  $B_h$  instead of increasing purchases from the market sector.

But how can investment add to aggregate profits? A firm invests today in order to obtain profits tomorrow, but how can aggregate investment today generate aggregate profits today? First, by investment a firm generates effective demand (revenues) for other firms. Second, investment does not increase costs for current but future output.

On the other hand, according to (8),  $I = \Pi - D + B_m - S_m$ , suggesting that investment is determined by profits, not the other way around. But this relation shows how investment is *financed*, not how it is *generated*. And the distinction between generation and financing is important. The generation of investment is discussed in more detail in Section 7 below, while, according to Section 2 above, investment is financed by retained profits ( $\Pi - D$ ) and net borrowings ( $B_m - S_m$ ) in the market sector – which are related to net borrowings in the other sectors through (13).

Of course, an increase of investment by  $\Delta I$  will not necessarily increase aggregate profits by  $\Delta I$ . If, for example, investment goods are imported, then  $X - Z$  in (14) will go down, moderating the positive effect of increasing investment on profits.

Note that wages in the market sector ( $W_m$ ) are not only costs for the market sector but also part of its effective demand. Even if profits depend negatively on wages both for a firm and a firm’s industry, this relation cannot be generalized to the whole market sector. Thus, paradoxically enough, aggregate wages in the market sector is not one of the ultimate determinants of aggregate profits.

We conclude that aggregate profits are determined by investment, dividends (including bonuses), export surplus, and excess borrowings by households and the government, as suggested already by Kalecki (1971 ch. 7). Without excess borrowings aggregate profits are consequently determined mainly by aggregate investment in the market sector, as emphasized

by Minsky (2008 p. 161). And investment today is made in order to generate profits tomorrow, while future profits depend on future investment, as also suggested by Minsky (2008 p. 163). The profitability of investment today consequently depends on aggregate investment tomorrow, making the basic engine in a market economy – the pursuit of profit – dependent on investment in real capital.

It also follows from (14) that a current account surplus, as well as borrowings in excess of savings by households and public institutions, can sustain aggregate profits if aggregate investment declines, as so often in recent years in many countries. But if household debt increases “too much” it can trigger not only inflation as usually measured but also, or alternatively, an asset-price bubble which, after bursting, may be followed by a prolonged recession as households reduce their “debt overhang”, as discussed, for example, by Mian and Sufi (2014) and Turner (2016). And even if a country’s current-account surplus can persist for a long time, particularly if its exporters face an expanding world market, it cannot persist unless there is at least one country with a corresponding current-account deficit, for example a country whose currency is treated as a reserve currency in international trade.

Moreover, if a government’s budget deficits are financed by selling bonds to households or firms, then increasing public debt may sooner or later imply a prohibitive burden of interest payments, or simply a refusal by the public to buy additional debt. However, if budget deficits are financed by selling government bonds to the central bank, then interest is paid by the government to the central bank and hence ultimately to itself. In this case the only restriction on increasing public debt is its effect on inflation.

Note that (14) can alternatively be written as

$$(15) \quad X - Z = (S_h - B_h) + (\Pi - D) + (S_p - B_p) - I,$$

where  $S_h - B_h$  is net saving by households,  $\Pi - D$  retained profits by firms, and  $(S_p - B_p)$  saving by public institutions (the budget surplus), so that the right hand side measures saving minus investment if we interpret  $\Pi - D$  as saving by firms. However, as emphasized by Blanchard (2009 p. 434), a trade surplus is, of course, not *determined* by saving minus investment. Eq. (15) is a *relation* between trade surplus, saving and investment which says, for example, that a country with a high saving rate (private plus public) must have either a high investment rate or a large trade surplus. But it does not expose the determinants of a trade surplus, which instead include the exchange rate, the demand for exports, the demand for imports, and their determinants.

## 2.5 Gross domestic product

The contribution of the market sector and the public sector to the output of final goods and services is usually measured by gross domestic product (GDP) defined as

$$(16) \quad Y = C + I + G + W_p + X - Z.$$

Note that I here distinguish between purchases by the public sector of goods and services from the market sector ( $G$ ) and purchases of labour from the household sector ( $W_p$ ). Note also that investment only includes spending by firms on real capital in the market sector, while purchases of new houses or apartments and other consumer durables are included in consumption (defined as spending by households).

GDP can be measured not only by expenditures, as in (16), but also by incomes. To prove this formally, we substitute the following expressions according to (9), (8), (10), (12) and (13) in (16):

$$C = W_m + W_p - T_h + F + D + A + B_h - S_h,$$

$$I = \Pi - D + B_m - S_m,$$

$$G = T_m + T_h - W_p - A + B_p - S_p,$$

$$X - Z = (S_m - B_m) + (S_h - B_h) + (S_p - B_p).$$

It follows that GDP also can be written as

$$(17) \quad Y = W_m + W_p + \Pi + F + T_m,$$

which is the sum of labour income ( $W_m + W_p$ ), capital income ( $\Pi + F$ ), and “public income” or, more precisely, taxes collected from the market sector ( $T_m$ ). Recall that  $F$  denotes interest payments by firms for borrowings from households, including interest on corporate bonds, while profits  $\Pi$  are distributed as dividends and bonuses to households and retained profits to firms. Note that income includes not only labour income and capital income but also taxes collected from firms, as also emphasized, for example, by Blanchard (2009 p. 44).

Thus, spending on final goods ( $Y$ ) equals income. However, this fundamental equality does not mean that all spending is financed by current income, or that all income generated in production is used to purchase all output. Part of the income will, of course, finance current spending, but part of it will be saved. And some spending will be financed by dissaving or borrowing, where some borrowing may be associated with an increase of the money supply, that is, the creation of money by commercial banks or the government.

GDP is an informative measure of final output in the “formal” sector, including both the market sector and the public sector but excluding home production and other production in the “informal” sector. However, national accounts include not only *nominal* GDP defined as a sum of expenditures at current prices on final outputs, as in (16), but also *real* GDP obtained by deflating nominal sales by price indices which include adjustment for quality changes and new products (as elaborated in Chapter 3).

Even if growth of real GDP, or of real GDP per capita, is an important indicator of the growth of output or living standards, it is a very blunt measure of a phenomenon with several dimensions.<sup>11</sup> Hence, in my attempts to explain the workings of a market economy I will not start from real GDP but from nominal GDP and its components – and the determinants of its components. Note also that a component of GDP can be measured by relating it to GDP, for example by  $C/Y$  for consumption, in the same way as government debt is usually measured, as per cent of nominal GDP. In the next two sections I will discuss in more detail the determinants of two components of GDP, namely consumption and investment, and in particular the possibility to write a component as a function of a small number of variables.

## 2.6 Consumption functions

According to (9) we have

$$(18) \quad C = W_m + W_p - T_h + F + D + A + B_h - S_h,$$

which shows how aggregate consumption in general depends on labour income  $W_m + W_p$  (including wages in firms producing investment goods or export goods) and capital income  $F + D$  but also on financial assistance  $A$ , borrowing  $B_h$  and saving  $S_h$ . Taxes are usually a function of income, e.g.  $T_h = t(W_m + W_p)$ , and in order to explain past consumption and predict future consumption we can specify and estimate a consumption function by relating borrowings and savings to incomes, explicitly or implicitly.

For example, by writing

$$(19) \quad C = c_0 + c_1(1-t)(W_m + W_p) + c_2D + A, \text{ where } 0 < c_1 < 1 \text{ and } 0 < c_2 < 1,$$

we assume implicitly that some wages are saved or used to repay debt, but also explicitly that all unemployment benefits and pensions etc ( $A$ ) are spent on consumption. This specification also implicitly assumes that all of households’ interest income ( $F$ ) is saved or used to repay

---

<sup>11</sup> On the distinction between the development of real GDP per capita and the development of living standards in a wider sense see, for example, Gordon (2016).

debt, while it explicitly assumes that some dividends and bonuses ( $D$ ) are spent on consumption, but with a marginal propensity to consume ( $c_2$ ) which may be different from the marginal propensity to consume out of labour income after taxes ( $c_1$ ). And if  $c_0 > 0$  this means that reducing wages by 1% will reduce spending on consumption goods by less than 1% even if  $D = A = 0$ , implying implicitly that dissaving or borrowing will smooth spending on consumer goods when wages vary.

Note finally that consumption depends on other components of GDP, like investment and exports, since wages in the market sector include wages for workers producing these types of goods. This creates an important interdependence between the components of GDP, as we shall see in more detail in Chapter 4.

## 2.7 Investment functions

The determinants of investment differ radically from the determinants of consumption. According to (8) investment can be *financed* by a mix of current retained profits ( $\Pi - D$ ), dissaving ( $-S_m$ ) based on past retained profits, and borrowing ( $B_m$ ). However, while a large household income may be associated with a lavish life style, large retained profits are not necessarily associated with large investments in real capital. In fact the contrary may hold, with large dividends or retained profits and low investment in maturing or declining industries, and low retained profits but large investment financed by large borrowings (perhaps based on the creation of money by commercial banks) in new and expanding industries. And since some investment may be restricted by credit, interest rates may also affect it.

Investment in an industry, both its type and its financing, will in general depend on the phase of the industry's product cycle: innovation, growth, maturity or decline. In the first two phases investment in real capital probably dominates, and financing may depend strongly on new equity or debt. On the other hand, dividends or savings, including investment in securities, may dominate for firms in the last two phases and in particular in a declining industry, where, in addition, investment in cost-reducing technology may occur.

Investment in additional capacity (buildings or machines) by existing firms should be particularly attractive when firms are producing at full capacity and demand is very elastic (so that sales increase so much that profits also increase with a larger capacity and a lower market-clearing price). Investment in inventories may be accidental (when production precedes sales and sales fall short of expectations) or intentional (for example when future

production costs are expected to rise). Investment in the development of new goods or new technology (R&D) is often characterized more by employment of researchers or engineers than by purchases from other firms. Starting a firm may be particularly attractive in new and expanding industries, where a new firm may obtain a large market share merely by entering the market at the existing market price.

Investment in real capital can be very variable over time (as measured by  $I/Y$ ) but also between industries (e.g. the consumer industry and the export industry) and between types (infrastructure, buildings, equipment, inventories, and R&D). It follows that the determinants of aggregate investment are so many and so diverse that they are difficult to summarize with a few variables, like current interest rates or profits. And to characterize the *change* of investment between years it may be necessary to report not only change of volume in nominal and real terms but also changes of type and industry.

The level of aggregate investment depends on an economy's relation to more advanced economies and its possibility to "catch up" by copying the best technology. In this case profitable investments may be "obvious" and an entrepreneur's problem is the supply of investment goods, not the demand. The growth of investment can be particularly high if employment in construction and other investment industries can grow rapidly at the same time as the consumer industry has sufficient capacity to produce basic consumer goods for a growing market employment without increasing inflation. This presupposes large labour reserves in a traditional subsistence economy, from which labour can be recruited at low wages. It also presupposes a financial system which can finance investment not only by savings but also by creating money.

On the other hand, in a rich economy on the technology frontier it may be difficult to perceive and realize profitable new investments, particularly if they involve "creative destruction", i.e. replacing an old industry or old methods by new ones. For people with money the demand for new investment in real capital may be much less obvious in an advanced economy, particularly since investment in financial instruments may be a tempting alternative. And people without money need a great deal of "animal spirits" to convince banks to finance their new and often fanciful projects.

Debt-financed investment in real capital is particularly important in developing countries, without accumulated profits as a source for financing. In such countries the demand for loans for real investment may be almost limitless, and then lending by money creation is particularly ingenious, since lending does not only create money but also production and income which can repay the loans.

In developed countries, on the other hand, most investment is financed by accumulated profits,<sup>12</sup> suggesting that nowadays most bank lending in developed economies is *not* for investment in real capital which generate production and income which can repay the loans. Hence the stability of a commercial bank will increasingly depend on the volumes and risks of direct and indirect lending for other purposes than investment in real capital.

## 2.8 Economic growth

Output in the long run – or more precisely the development of output as measured by GDP at constant prices – is determined by the history of its growth, that is, by the accumulation over time of yearly changes. Thus, a theory of economic growth must *ultimately* be able to explain a change of output from one year to the next. The first step is a decomposition of a change of nominal output according to (16) or nominal income according to (17), that is,

$$(20) \quad dY = dC + dI + dG + dW_p + dX - dZ, \text{ or}$$

$$(21) \quad dY = dW_m + dW_p + d\Pi + dF + dT_m,$$

where  $dY$  is a change of  $Y$  between two adjacent years. The corresponding changes in real terms require adjustment with a price index, perhaps a consumer price index for all terms in (21), but a price index for investment goods for the second term in (20) etc. Note that some of the terms in (20) or (21) may be positive and some negative.

Moreover, each term in (20) can be decomposed according to industry, reflecting structural changes, and including changes in the distribution of effective demand (spending) between industries. For each industry the change in output ( $q$ ) can then be related to the change in employment ( $N$ ), so that a change in labour productivity ( $a$ ) can be estimated as

$$(23) \quad da/a = dq/q - dN/N.$$

And then the problem is to explain an increase in labour productivity in an industry by relating it to recent investments in real capital or changes of skills, inputs or organization.

### *Approximations*

Of course, output as measured by GDP at constant prices ( $\bar{Y}$ ) should be correlated with real capital ( $K$ ) and labour ( $L$ ) as measured in national accounts, since more factories and machines as well as more workers should imply more output. This association suggests that

$$(24) \quad \bar{Y} = AK^\alpha L^\beta$$

---

<sup>12</sup> See, for example, Cecchetti and Schoenholtz (2011 p. 322).

might be an approximation of the relation between output and its determinants, even with  $\beta = 1 - \alpha$ . However, this approach has not been capable of explaining large differences between countries, or the development over long periods of time in one country.<sup>13</sup>

On the other hand, (24) also suggests a more general approach, namely explaining output by a small number of variables, including not only capital and labour but also, for example, measures of oil (or energy), education, innovation, R&D activity, infrastructure, and institutions, as modern theories of economic growth illustrate. Perhaps measures of free trade and, more generally, a measure of specialization or “division of labour” should be added to this list. In fact, a theory of economic growth is ultimately a theory of economic development, which in general should be assisted by narratives in terms of institutions, economic policy, and historical accidents.<sup>14</sup>

Moreover, additional output is caused not only by additional capital or additional labour, or by new technology or new institutions, but also by increased spending. Periods of increased spending without increased capacity may even be a necessary component of economic growth “in the long run”, namely if investment in additional capacity is not undertaken in an industry unless the industry is producing at full capacity.

## 2.9 Relation to the literature

The fundamental identity is the centrepiece of “macroeconomic accounting” in Wray (2012 ch. 1), but is only suggested in traditional textbooks. For example, Blanchard (2009 p. 434) notes that a relation close to the fundamental identity and involving net exports must be interpreted as an identity and not as an equation exposing the determinants of net exports.

The profit equation suggested by Kalecki (1971 ch. 7) and Minsky (2008 ch. 7) is not even mentioned in traditional textbooks. It may be that the proofs given by Kalecki and Minsky are so sketchy that they are hard to believe. On the other hand, the relation has not been disproved by orthodox economists, only ignored. Perhaps the formal proof in this chapter can make a difference.

---

<sup>13</sup> See, for example, (Jones 2014).

<sup>14</sup> For introductions to economic growth and economic development see, for example, Gordon (2016), Acemoglu and Robinson (2013), and Marks (2015).

## Chapter 3: Inflation

Sellers will not accept money as a means of payment unless money also is a store of value, which presupposes an acceptable amount of inflation. But the purchasing power of money – the value of money in terms of goods – depends on the prices of what buyers intend to buy, so the concept is ambiguous, as emphasized already by Knut Wicksell.<sup>15</sup> Different agents may be interested in different goods and hence also in different prices and in different summary measures, even if a price index summarizing the development of prices of basic consumer goods is important for most people.<sup>16</sup>

Since inflation means increasing prices, Section 1 deals with price formation and in particular the determinants of price changes, including changes in labour costs, costs of other variable inputs, mark-ups and labour productivity. The relation between inflation and money growth is also discussed. Section 2 defines inflation as a stochastic variable and notes that so far only weighted averages are reported in official statistics. It also reviews some of the most important price indices and argues that information on inflation should include not only its distribution but also its determinants. Section 3 on monetary policy discusses the choice of inflation target and how a central bank can control inflation by controlling effective demand through its policy rate. Restrictions on the choice of policy rate are discussed in Section 4, particularly restrictions in an open economy. The last section relates the text to the literature.

### 3.1 Pricing and inflation

There are two possible “regimes” for price formation in an industry, and which regime rules depends on the relation between effective demand and capacity, as elaborated in Chapter 1. Of course, the regime is not necessarily the same in all industries: while some industries may have production restricted by capacity and prices determined by market clearing and effective demand, other industries may have production restricted by sales at prices set by firms.

Thus, price increases in an industry depend entirely and directly on demand if all firms are producing at full capacity. Higher prices will in general be associated with higher production, but production can be only marginally higher, assuming that the supply curve is steep even if it is not vertical. On the other hand, if production is restricted by sales at prices set by firms,

---

<sup>15</sup> See, for example, Wicksell (2010 [1906] ch. IV).

<sup>16</sup> However, a bundle of “basic consumer goods” cannot be representative of all households, independent of income, so perhaps there should be a consumer price index for poor people and another consumer price index for rich people, as suggested in Stiglitz, Sen and Fitoussi (2010 p. 48).

then inflation is determined by firms when they revise their prices according to mark-up pricing,  $p = (1 + m)c$ , where  $c$  is direct cost (variable costs per unit of output) and  $m$  is the mark-up on direct costs set by a firm if it is a price maker (price leader) and determined by the market price if it is a price taker.

Assuming in addition that employment ( $N$ ) and other inputs are proportional to output ( $q$ ), we have seen – in Section 5 in Chapter 1 – that the determinants of direct costs include labour productivity ( $a = q/N$ ), the wage rate ( $w$ ) and variable costs other than labour costs measured as a share of labour costs ( $h$ ), or more precisely that

$$(1) \quad p = (1 + m)(1 + h)(w/a) \text{ and } q = D(p) \text{ if } D(p) < k,$$

where  $k$  denotes the firm's capacity and  $D(p)$  its sales at the price it sets ( $p$ ).

Thus, when production is restricted by sales, inflation depends on the evolution of labour costs, costs of other variable inputs, mark-ups and labour productivity. Production depends entirely on demand for the firm's products at the new prices, while employment depends on production and labour productivity. In this case higher prices in an industry will not be associated with higher but with lower production, unless demand has increased.

Inflation is usually measured by a consumer price index (CPI). An increasing consumer price index presupposes that a weighted average of prices on consumer goods is increasing, which presupposes price increases in at least some consumer industries. The first step in explaining inflation is consequently to relate it to inflation in different industries. And the second step is to explain inflation in the various industries. The explanation is not necessarily the same in all industries. In some consumer industries inflation may depend on increasing demand when firms are producing at full capacity, while inflation in other industries may depend on increases in import prices or wages increasing more than labour productivity.

Inflation is correlated with money growth. But money growth can never be the cause of inflation even if it sometimes may be associated with the cause of inflation, namely excessive effective demand caused either by excessive lending and money creation by commercial banks or excessive government spending financed by selling bonds to the central bank. Normally, however, inflation is *caused* by price formation and *accommodated* by money growth generated by new lending by commercial banks supported by a central bank which supplies reserves on demand to the payments system.

More precisely, lending by commercial banks always increases the stock of money and may, when borrowings are spent, also increase the effective demand for consumer goods. But this affects inflation only when production is restricted by capacity, since when production is

restricted by sales, effective demand affects production and employment but not prices. Moreover, even if an increasing stock of money may precede inflation, inflation may also precede money growth, since higher prices may increase borrowings in monetary terms.

As a case study, let us try to explain the development of inflation since World War II in the U.S., where inflation was first increasing, from less than 2 per cent per year in 1960 to more than 10% in 1980, and then decreasing, from about 4% in the 1980s to about 2% since 2000.<sup>17</sup> A particularly interesting period is the 1970s, when inflation was not only high but also volatile, varying from 10% in 1974 to about 5% in 1976, 12% in 1980 and 3% in 1983. Each of the two large price increases in the 1970s was associated not only with a sharp price increase in oil (the first one in 1973 to 1975 and the second in 1979 to 1981) but also with an increase in unemployment, a combination of high inflation and high unemployment which since then has been called *stagflation*.<sup>18</sup>

Now, to explain inflation we first of all have to decompose it, a decomposition which shows how a change in CPI depends on changes of effective demand (in industries where production is restricted by capacity) and how it depends on changes in wages, input prices, labour productivity and mark-ups (in industries where production is restricted by sales). But a decomposition of CPI in the U.S. is outside the scope of this text.

Of course, a plausible hypothesis is that the increasing price of oil, which is an important input in all production, either directly or indirectly, is an important part of the story. Moreover, some of the increasing inflation since the 1960s may be due to increasing employment in the public sector, which increases the effective demand for consumer goods without increasing their production. The Vietnam War must have added to this pressure. On the other hand, increasing labour productivity in firms producing or distributing consumer goods probably moderated the effect of increasing wages on inflation during the 1960s and 1970s.

### 3.2 Defining and measuring inflation

Prices of final products change as market conditions change: some decrease steadily over time because of technological improvements, some always increase and some fluctuate, with or without a trend. Inflation is consequently a variable which depends on the product. This means that we can interpret inflation as a stochastic variable  $X$  with a probability distribution defined by assigning to the event  $a \leq X \leq b$  the “relative importance” of those products which

---

<sup>17</sup> See, for example, Blanchard (2009 p. 190).

<sup>18</sup> See, for example, Blanchard (2009 pp. 173, 177).

have a price change between  $a$  and  $b$ , where the relative importance is measured by the monetary value of the output of the relevant products divided by nominal GDP.

The mean of  $X$  is an unambiguous measure of the change in the value of money in terms of goods only if the standard deviation is sufficiently small, since only then is inflation (almost) independent of the goods a buyer wants to purchase. Hence it is unfortunate that so far only means (weighted averages) of inflation are reported in official statistics, particularly since data collection often already includes information on the distribution.

Inflation may differ substantially between types of goods, and it is important to distinguish between consumer goods, investment goods, and export goods. For example, Statistics Sweden publishes price indices for consumer goods, new dwellings, import goods, export goods, intermediate goods, and services. Price indices for main groups of consumer goods are also published, including food and non-alcoholic beverages, clothing and footwear, housing (including water, electricity and fuels), furnishing, and household goods, etc.

Information on inflation for consumer goods is usually based on a “basket” of representative consumer goods. Observing prices for the goods selected, often every month, and attributing weights to the observations made, average inflation is calculated, usually based on an index of the price of the basket. However, the consumer goods in the basket may not be representative of all consumer goods, and the weights may not be the same as defined above. Only “basic” consumer goods may be included and the weights may reflect the spending of a “typical” consumer instead of all consumers’ spending.

In practice the measurement of a price index is complicated by both changes in the set of goods produced in an economy and changes in the quality of produced goods. Changes in the set of representative goods in a basket are in practice handled by updating the set of representative goods between years and “chaining” prices of the baskets for adjacent years, while a quality-adjusted price is an observed price reduced by  $x$  per cent if the increase in quality is estimated to be  $x$  per cent.

However, estimation of quality changes are often based on subjective factors, for example an interviewer’s estimation of a buyer’s evaluation in monetary terms of quality changes in home electronics or cars. And if quality-adjusted price changes are so large that they have a substantial impact on a price index, for instance by reducing the increase of a consumer price index from 2 to 1 per cent, then the methods used for quality adjustment becomes particularly important. Perhaps a price index should be reported both with and without quality-adjusted prices if quality-adjustment matters for the measurement of inflation. Inflation on services is

measured in the same way as inflation for goods even if quality-adjustment may be more difficult (or even impossible) in some cases, like consulting and banking.

A measure of the “price level” for all final goods produced in an economy is the GDP deflator, defined as the ratio of nominal GDP to real GDP, where real GDP for a year is defined as quantities of final goods produced during the year multiplied by prices in a base year, at least in principle.<sup>19</sup> The GDP deflator can consequently be interpreted as the price level expressed as a percentage of the price level in a base year, and the relative change between two years of this index is a measure of the average inflation rate. But note that this is an indirect approach, which only can be used to measure average inflation, not its standard deviation.

Information on inflation should include not only its distribution (standard deviation) but also its determinants. Suppose, for example, that production is restricted by capacity in industries producing and distributing consumer goods. In this case inflation depends on the growth of effective demand for consumer goods and can only be reduced in the short run by reducing the effective demand for consumer goods.

On the other hand, if production in a consumer industry is restricted by sales and not capacity at prices set by firms, then price increases will, according to equation (1), depend on wage increases or increases of input prices or mark-ups. If inflation is due to higher input prices or higher mark-ups, then higher prices will reduce sales and production and may also reduce employment and increase unemployment. If instead inflation is due to wage increases, then the higher wages may, unless they are saved, increase effective demand so much that the effect of higher prices on sales may be counteracted by higher effective demand. In this situation reducing effective demand for consumer goods by economic policy may, of course, completely neutralize the effect of higher wages on consumption, but it cannot reduce inflation unless wage formation is affected.

It follows that information on the determinants of inflation as measured by a consumer price index should include the following aspects of the consumer industry or its parts (each part corresponding to a component of the consumer price index): inflation regime (market clearing or mark-up pricing), labour productivity, wages, prices on raw materials and other imports, and mark-ups. A first step towards information on the determinants of inflation are measures of “underlying” inflation contemplated by central banks, since such measures are obtained, for example, by eliminating changes in interest rates on housing costs or indirect

---

<sup>19</sup> In practice, when the set of final goods is renewed between years, the index is a chain index.

taxes or temporary price changes from a monthly consumer price index. But a more transparent way of reporting such information would be a decomposition of consumer price inflation which shows the contribution of changes in interest rates and indirect taxes but also changes in prices considered to be volatile.

### 3.3 Monetary policy

Policy makers do not normally attempt to stabilize individual prices. But they attempt to stabilize the purchasing power of money, which in practice means stabilizing a consumer price index, or more precisely its rate of increase. A common view is that inflation should be above all stable but also low, and a common target for central bankers is two per cent per year. But how can this level be motivated?

#### *Choosing a target for inflation*

It is not easy to define an “optimal” inflation rate as measured by a consumer price index. Two per cent appears to be a common choice by central banks, and the motivation often appears to be that even if 0 per cent is the ideal, it cannot be reached without increasing the risk for deflation too much.<sup>20</sup>

But is deflation always bad? Consumers usually react positively to falling prices on consumer goods, so deflation *per se* cannot be bad. It is sometimes argued that falling prices may lead to consumers postponing their purchases, but even if this may apply to some households searching for a house to buy, or some consumers waiting for a bargain sale to take place, the effect of marginally falling prices on the timing of sales and production can only be marginal. Note also that if an increase of a consumer price index is about 2% per year, then the weighted average of price changes is so close to 0 that many prices are declining even in normal times.

Of course, if falling prices are due to falling wages – and not increasing labour productivity or decreasing import prices or falling mark-ups – then the real value of debt in terms of wages will increase and may even cause bankruptcy for some households. But in this case the real problem is not falling prices but falling wages.

Can price deflation without wage deflation be a problem? While decreasing prices at given wages are good for consumers they can be bad for producers, by reducing their profits,

---

<sup>20</sup> However, in the literature “inflation targeting” can also mean not only a target of *at most x%* but also a target of *at least x%*, the motivation being that, when anticipating higher prices in the future, consumers will add to effective demand by spending their money now instead of later.

and ultimately also bad for consumers, by reducing their employment and earnings. First, if production is restricted by capacity in an industry and the market price consequently determined by market clearing, then a lower market price – which always is caused by lower effective demand – will also reduce profits. But as long as firms are producing at full capacity, this price decline will only marginally reduce production and employment.

Second, if production is restricted by sales at prices set by firms in industries producing consumer goods, then falling prices without wage deflation can only be due to increasing labour productivity or falling input prices or lower mark-ups. In this case deflation may even be associated with *increasing* production and employment, other things being equal. Of course, a fall in effective demand during a recession may reduce sales so much that some firms panic and try to counteract a decrease in individual sales by attempts to increase market shares by price cutting, even if keeping the market price unchanged will maximize profits (or minimize losses) even in the new situation. But even without irrational price cutting, a fall in effective demand during a recession may reduce sales and also profits. And to counteract a decline in profits, firms may oppose all wage claims by workers and even attempt to reduce wages, thus contributing to wage deflation.

Thus, price deflation is problematic if – and only if – it is caused by wage deflation, so an inflation rate below 2% per year is not necessarily problematic. On the other hand, what about an inflation rate which is somewhat higher than 2% per year, say 4%? Inflation is certainly bad for senior citizens with fixed nominal pensions and for unemployed persons with fixed ceilings for benefits in nominal terms. It is also bad for classical rentiers, with fixed income from bonds. On the other hand, inflation is not necessarily bad for people with real estate and equities, since the market value of such assets usually follow inflation. Households with debt in nominal terms will even benefit from inflation, provided that their earnings follow inflation. Since government debt as a share of GDP will fall more rapidly with inflation, moderate inflation is also a smooth way of reducing the burden of government debt, as emphasized, for example, by Krugman (2012 p. 141) for the U.S. after World War II.

What about deviations from an inflation target caused by large increases of the price of oil or other imports? Such deviations can only be neutralized by compensating decreases of wages in the consumer industry (when production is restricted by sales and not capacity at prices set by firms). Should policy makers attempt to enforce wage moderation or should they accept a temporary deviation from the target in this case, perhaps even anticipating that a decline of import prices in the immediate future will reduce inflation below the target and make *average* inflation close to the target? It could be argued that *real* wages are not affected

by the choice, making a “strict” target attractive. But if wage moderation cannot be enforced without costs in the form of unemployment, then a “flexible” target would be more attractive.

### *Transmission mechanism*

The basic idea of modern monetary policy is that a central bank can control inflation by controlling effective demand through its policy rate. Of course, a central bank can always restrict effective demand by a sufficiently high policy rate. But will this also reduce inflation? A central bank can also stimulate effective demand by lowering the policy rate, even if sometimes not even a policy rate equal to zero may be a sufficient stimulus. But will this also raise inflation? Let us first discuss the relation between the policy rate and effective demand.

Since the policy rate affects the level of all other interest rates, it can also influence effective demand. And the effects are transmitted through three channels: 1) by affecting new borrowings; 2) by affecting interest payments for existing debt; and 3) by affecting asset prices and hence the wealth of economic agents so much that it also affects their expenditures.

A sufficiently high interest rate will reduce effective demand through all of these channels. First, it may even, if sufficiently large, completely eliminate borrowings needed for investment by firms (in buildings and machinery) and investment by households (in new homes and cars). Second, by increasing payments for existing debt (with variable interest rates) a higher policy rate will reduce disposable incomes for households with debt and profits for firms with debt, perhaps even so much that some households and firms go bankrupt. (On the other hand it will also *increase* disposable income for households and firms with savings, but the effect on expenditures of this increase may be relatively small.) Third, increasing interest rates tend to reduce prices for homes and securities and hence the wealth and probably also the expenditures of the owners of such assets.

On the other hand, lower interest rates should raise effective demand through all of these channels, even if the effects may be small in practice compared to the effects of all other determinants of effective demand, particularly at low rates. For example, firms’ investment in capacity and new technology depends on expectations of future profitability of investment in real capital relative to investment in other assets like securities. And expectations of profitability in an industry depend on the phase of the industry’s product cycle (innovation, growth, maturity or decline) and on expectations of sales and costs in the future. In developed economies most investment in real capital is financed by internal funds,<sup>21</sup> which can make

---

<sup>21</sup> See, for example, Cecchetti and Schoenholtz (2011 p. 322).

expected profitability almost independent of borrowing costs for many firms. On the other hand, expectations of *relative* profitability also depend on expected profitability of investment in securities, which depends on the policy rate.

Thus, a higher policy rate may reduce investment in real capital mainly because it encourages businesses with internal funds to invest in securities instead of real capital. And a lower policy rate may stimulate investment in real capital mainly because it reduces borrowing costs for innovative and expanding businesses without internal funds – provided that banks are willing to lend money to such firms.

However, low interest rates do not necessarily stimulate investment in real capital. Some financial institutions may be unwilling to lend by buying corporate bonds at low interest rates because they find it more profitable to have money ready for buying bonds at higher interest rates in the future, as emphasized by Keynes (1936). But commercial banks mainly lend by creating money, not by investing a given stock of money, so a more plausible explanation for banks' unwillingness to lend during a recession with a low policy rate is that they find proposed investment projects too risky, or that they have to repair their balance sheets (reach a required capital ratio) after some heavy losses.

On the other hand, businesses may be unwilling to borrow during a recession – even at very low rates – as emphasized by Koo (2008). Businesses may be particularly unwilling to borrow after the burst of an asset-price bubble – especially if they have invested heavily with borrowed money in real estate or stocks before the burst of the bubble and have to repair their balance sheets. For other businesses the reason for not wanting to borrow at low interest rates during a recession may simply be that they for the time being find the future uncertain and investment in real capital too risky, particularly investment financed with borrowed money.

The policy rate may also affect consumption. A low policy rate may even trigger an asset-price bubble and increased consumption financed by borrowing based on increasing market values of homes. On the other hand, after the burst of an asset-price bubble not even a low interest rate may be sufficient to prevent households from increasing savings or reducing debt.

The main tool of modern monetary policy is the policy rate, but intervention in securities markets by a CB can also be used to supply liquidity and reduce long-term interest rates. Such measures, often called “quantitative easing”, have also been used extensively during and after the global financial crisis in 2007-2009. For example, by purchasing securities in secondary markets a CB increases the reserves of commercial banks and – if the sellers are other institutions than commercial banks – the stock of money. Increased liquidity (i.e. reserves) may also make banks more willing to lend, and reduced long-term interest rates will make it

easier for large firms to finance investment by selling bonds. The purchase of hard-to-value securities may be particularly important in a credit crunch, to facilitate new lending.

*Effective demand and inflation when production is restricted by sales*

It remains to see how a central bank (CB) by acting on effective demand through its policy rate can affect inflation. Inflation as usually measured is determined by the pricing of consumer goods. We have seen that pricing depends on the relation between capacity and effective demand, and that there are two possible regimes: *mark-up pricing* if production is restricted by sales and *market clearing* if production is restricted by capacity.

In the first case, when effective demand for consumer goods is sufficiently small at current capacity, sales and production will be less than the capacity of the consumer-goods industry at the prices set by firms according to equation (1). Prices consequently depend on wage rates, input prices, labour productivity, and mark-ups. Unless mark-ups change, inflation must then be due to wages increasing more than labour productivity or to increasing prices on imports or other intermediate goods.

Of course, a central bank cannot affect labour productivity. It can affect import prices by affecting the exchange rate, which it can affect in the long run by affecting domestic inflation relative to foreign inflation, and in the short run by intervening in the foreign exchange market or setting a policy rate which differs from foreign policy rates. However, in a system with flexible exchange rates interventions are not necessarily carried out, even if they by stabilizing a volatile exchange rate can stabilize import prices. A policy rate above foreign rates may result in an inflow of capital which appreciates the currency and hence moderates price increases on commodities and other imports. However, in a world with free mobility of capital, interest rates cannot deviate much, as argued in Section 4 below.

How can a CB affect wages? It cannot affect wage formation directly. Hence a CB cannot control inflation in a mark-up regime unless it can control wage revisions indirectly, by creating and using unemployment as a disciplinary device when inflation is above the target, and reducing unemployment in order to increase wages when inflation is below the target. In practice, however, a revision of the wage level in the consumer-goods industry is the result of a much more complicated process, sometimes involving negotiations between employers and employees or their organizations, as discussed in more detail in Chapter 6. It may even be exogenously determined, for example by a revision of the wage level in the export industry, accepted as a norm by all parties in the consumer industry.

Thus, the crucial determinant of wage revisions is not unemployment but a *wage policy* determined by the interaction between employers and employees or their organizations within a framework of laws enacted by the parliament and regulations issued by the government. Increasing effective demand may then increase production and employment in the consumer-goods industry without necessarily raising inflation because of higher wages caused by lower unemployment. And this is true even if wages are determined by trade unions, since rational unions will realize that prices are adjusted to wages and adjust their wage claims to this fact.<sup>22</sup> Moreover, if inflation increases above its target because of increasing import prices, it would be bad policy to try to reduce inflation by reducing effective demand, since this would only reduce production and employment and raise unemployment without affecting inflation.

If inflation is too high when production in the consumer industry is restricted by sales at prices set by firms, and if the main cause of high price inflation is high wage inflation, then a CB should first of all ask why wage inflation in the consumer industry is so high, since this is somewhat of a mystery. To see why this is puzzling, note that it follows from (1) that

$$(2) \quad w/p = \frac{a}{(1+h)(1+m)},$$

where  $a$  denotes labour productivity,  $h$  measures the cost of intermediate goods as a share of labour costs, and  $m$  is a mark-up on direct costs. It follows that wage setting in the consumer industry cannot influence real wages ( $w/p$ ) for its workers, making excessive wage claims by unions hard to understand. Wage drift caused by competition for workers may be an explanation, at least when excess capacity is combined with labour shortage.

Finally, *deflation* can only be explained by decreasing import prices or wages increasing less than labour productivity in the consumer industry. And then a CB is powerless, since it cannot raise wages by reducing the policy rate, unless this rate is high to begin with and reducing it will increase sales, production and employment so much that the concomitant decrease in unemployment also raises wage claims by workers and wage concessions by firms.

#### *Effective demand and inflation when production is restricted by capacity*

When effective demand for consumer goods is so large that production is restricted by capacity and not sales, then prices are market clearing and inflation depends entirely on the effective demand for consumer goods, while production – and employment – will be close to capacity and thus hardly influenced by effective demand. Unless wages are fixed in long-term

---

<sup>22</sup> Complications are discussed in Chapter 6.

contracts, there is also a risk for wage-price spirals, with increasing wages adding to increasing effective demand and more inflation, even if production at full capacity and increasing inflation may stimulate investment in new capacity and a return to mark-up pricing. And then a CB by reducing the effective demand for consumer goods can reduce inflation without reducing employment, at least not employment in the consumer industry. But note that increasing the policy rate will not only reduce the effective demand for consumer goods but also discourage investment in increasing capacity in the consumer industry.

Production and capacity in other industries will certainly matter for inflation as measured by a consumer-price index, but only to the extent that production, through the earnings it generates, affects effective demand for consumer goods. Thus, increasing production of export goods, investment goods or public goods (i.e. goods purchased by the public sector from the market sector) will increase the risk for inflation through the effect of increasing employment and earnings in these industries on the effective demand for consumer goods. Increasing employment in the public sector will also increase the risk for inflation.

We conclude that a CB sometimes can reduce inflation without raising unemployment (when production in the consumer industry is restricted by capacity), and sometimes reduce unemployment without raising inflation (when production in the consumer industry is restricted by sales at prices set by firms). It follows that a CB need not focus exclusively on inflation as a target.

### **3.4 Restrictions on the policy rate**

In an open economy the possibility of free capital flows between countries may equalize interest rates between countries, assuming that investors move their capital from markets with low yields to markets with higher yields. As also emphasized by Blanchard (2009 p. 456), if interest rates are equalized by perfect capital mobility, then a change in interest rates presupposes either coordination among central banks or leadership by one of them.

Leadership by the Federal Reserve means, more precisely, that its policy rate determines the *level* of policy rates set by other central banks. Even if some deviation is possible, this is, of course, an important restriction on a country's policy rate.

#### *Restrictions on interest rates in an open economy*

Suppose that the yield on bonds in the Swedish capital market is higher than the yield on bonds in the U.S. Then some investors in the U.S. will buy bonds in Sweden. Since this will raise bond prices in Sweden and consequently reduce their yields, while it hardly affects bond

prices in the U.S., it can be argued that the interest rate on bonds in Sweden will adjust to the American rate.

However, risk premiums and flexible exchange rates will complicate the picture. A difference in bond yields may be balanced by a risk premium and then there will be no capital flows which can equalize interest rates. And with flexible exchange rates an American investor in Swedish bonds may fear a depreciation of the Swedish *krona*, which would reduce the dollar value of the bond's dividends and price, and make the purchase of Swedish bonds less attractive.

What about short-term interest rates: will capital flows equalize interest rates in money markets? A central bank can determine the short-term interest rate in money markets by controlling the interbank rate of reserves. But what happens if the short-term interest rate in Sweden deviates "much" from the short-term interest rate in the US? Suppose, for example, that the Swedish money-market rate is higher than the American rate. Then some American investors may be tempted to invest in Swedish bills instead of American bills and obtain some extra interest income in dollars after a few weeks or months, provided the exchange rate has been stable.

An increasing demand for Swedish bills from American investors will, of course, increase the price and reduce the yield. On the other hand, to support its target for the interbank rate of reserves a CB can intervene directly in the money market. And a substantial gap between the policy rate and the money-market rate would be embarrassing for a CB by showing that it cannot control a basic interest rate in an economy.

However, if the Swedish money-market rate is higher than foreign rates, it will be tempting not only for foreign investors to buy Swedish bills, but also for domestic banks and firms to borrow in international money markets with lower rates. And then the private sector will accumulate debt denominated in foreign currencies which in a crisis with a large depreciation of the domestic currency may be impossible to repay for some banks or firms. Thus, to minimize the dangers associated with debt denominated in a foreign currency, a CB may find it wise to follow the lowest policy rate set by other central banks.

This argument suggests *price leadership* as a market form even for central banks, with the Federal Reserve as price leader. Thus, if the Fed reduces its policy rate, other central banks will follow suit, assuming they want to minimize borrowing in a foreign currency by domestic banks or firms. If the Fed raises its policy rate, other CBs may also follow suit, but for another reason. Higher interest rates in the American money market will attract short-term money from other countries if exchange rates are stable. Moreover, an outflow of capital from a

country tends to reduce the value of its currency and hence raise the profits from first buying American bills and obtaining a higher interest rate in dollars and then exchange dollars for domestic money. And anticipating such a development, investors may even increase their investments in the American money market, reinforcing the upward pressure on the domestic money-market rate.

Note finally that the relation between capital mobility and exchange rates is a problematic one. Volatile exchange rates may reduce capital flows because they increase the risk for losses. On the other hand, volatile exchange rates may stimulate speculative capital flows. And large capital flows may contribute to exchange rate volatility. However, with (approximately) equal and stable interest rates, flows of capital between countries should be dominated by the financing of trade in goods and investment in real capital. If this is not the case in the real world, it is tempting to conclude that the combination of free capital mobility and flexible exchange rates has created a system where both capital flows and exchange rates are unpredictable.

#### *Interest rates as objectives*

The interest rate, or more precisely the policy rate of a central bank, is usually not thought of as an objective but as an instrument of stabilization policy. However, the policy rate is an objective in the sense that it is a target for the interbank rate of reserves which a CB cannot achieve merely by announcing it. In fact the interbank rate of reserves can be very volatile in a market without intervention by a CB, or with a CB focused on stabilizing the growth of money supply, as shown by the fluctuations of the federal funds rate in the U.S. until 2002.<sup>23</sup> Thus, stabilization policy is in practice very much concerned with stabilizing the interbank rate of reserves, particularly in systems with a “corridor” for the interbank rate, implying not only a ceiling for the rate but also a floor.

However, the policy rate is usually thought of as an instrument for stabilizing inflation and output, through its effect on all other interest rates and hence indirectly on effective demand. With the policy rate perceived as a tool, it is only its effect on inflation and output which should matter. But the *variability* of the policy rate also matters for firms and households, and the policy rate has since the 1970s been very variable in many countries. Even if a CB often announces not only a policy rate but also a plan for its development over the near future, the variability of the policy rate makes planning of investment more difficult, both investment in

---

<sup>23</sup> See, for example, Cecchetti (2008 p. 432).

housing by households and investment in productive capacity and new technology by firms. In fact, “by keeping interest rates stable, policymakers can insulate the real economy from disturbances that arise in the financial system”, as noted by Cecchetti (2008 p. 503).

Of course, the *level* of the interest rate also matters for economic agents (as measured by its average over a business cycle), even if they cannot agree on what the level should be: while borrowers prefer low rates, lenders – including pension funds – prefer high rates. A complication is that preferences on the (nominal) interest rate depend on inflation, even if an inflation rate which is stable at about 2 per cent should make this problem a marginal one. Perhaps there exists a “critical value” of the level of the policy rate, such that both lenders and borrowers can accept it as “fair”, while a deviation from it will be opposed by either borrowers or lenders. In any case, we cannot *a priori* exclude preferences on the interest rate which depend on other effects than effects on inflation and output. However, postulating a target for the level of interest rates, the policy rate can no longer be thought of as an instrument for controlling inflation. And then, as discussed in Chapter 6, a government has to stabilize the value of money by other means than the policy rate.

### 3.5 Relation to the literature

According to the quantity theory of money, changes in the aggregate price level are caused solely by changes in the quantity of money, at least “in the long run”. The theory is based on the assumption that the value of an economy’s output is proportional to its stock of money,

$$(2a) \quad Py = VM ,$$

where  $P$  is the price level,  $y$  is output of final goods in real terms (GDP at constant prices),  $M$  is the stock of money, and the constant of proportionality  $V$  measures the “efficiency” – or “velocity” – of money. This equation also implies that

$$(2b) \quad dP/P = dM/M + dV/V - dy/y ,$$

so that, for example, if the velocity of money is constant ( $dV/V = 0$ ), and real GDP grows by 3 per cent per year ( $dy/y = 0.03$ ), then the inflation rate is 2 per cent per year ( $dP/P = 0.02$ ) if – and only if – the stock of money grows by 5 per cent per year ( $dM/M = 0.05$ ).

Equation (1) reflects the fact that money is both a unit of account (for prices) and a means of payment (for transactions), implying that the stock of money and the value of output must move together. The correlation between the growth of money and inflation is particularly

strong when inflation is high.<sup>24</sup> However, in this case it is also particularly clear that money growth and inflation have a common cause, for example a government which cannot finance its expenditures by taxes or borrowing and consequently has to “print money”, that is, sell bonds to its central bank (CB) and spend the proceeds, as in the former Soviet republics in the 1990s.<sup>25</sup>

Note, however, that high money growth and high inflation can also be due to excessive private spending financed by excessive lending by commercial banks if this is accommodated by the CB, that is, if the CB lends reserves to commercial banks on demand. Note also that higher spending will lead to inflation instead of more production only if production in the consumer industry is restricted by capacity instead of sales. And in countries with moderate inflation it is even more difficult to attribute inflation solely to money growth, because the velocity of money, as measured in practice by  $V = Py/M$ , is very volatile in the short run.<sup>26</sup>

### *The Phillips curve*

In the short run inflation is linked to the Phillips curve, originally a plot of the rate of *wage* inflation against the rate of *unemployment*, but nowadays a plot of the change in the rate of *price* inflation against either unemployment or output. Moreover, if high unemployment is associated with decreasing inflation, and low unemployment is associated with increasing inflation during a certain period, there is a certain rate of unemployment at which inflation is stable. This level of unemployment is called “equilibrium unemployment”, and if such an equilibrium level exists, the Phillips curve can be formulated as a plot of the *change* in the inflation rate against the *deviation* of the rate of unemployment from its equilibrium level. And then the equation of the Phillips curve can be written, for example, as

$$(3) \quad \pi_t - \pi_{t-1} = -\alpha(u_t - u_e),$$

where  $\pi_t$  denotes the rate of inflation in period  $t$ ,  $u_t$  the rate of unemployment in period  $t$  and  $u_e$  the equilibrium rate of unemployment.

Now, inflation must, of course, also be linked to pricing in a theory of inflation. And then *expected* inflation enters the picture in traditional economics. Thus, it is assumed that inflation depends not only on deviation from equilibrium unemployment but also on expected inflation.

---

<sup>24</sup> Cecchetti (2008 p. 483).

<sup>25</sup> Cecchetti (2008 p. 485).

<sup>26</sup> Cecchetti (2008 p. 491).

There are (at least) two schools in the literature, one where inflation expectations are held by price setters, and one where inflation expectations are held by wage setters.

*Inflation, unemployment, and inflation expectations by price setters*

Inflation is about price changes, so a theory of inflation must start from a theory of pricing. It is argued, for example, that “firms set the amount by which they raise their prices on the basis of their expectations of the economywide inflation rate and the state of demand for their products...”,<sup>27</sup> so that inflation can be written as

$$(4) \quad \pi_t = \pi_t^e - \alpha(y_t - y_e) + \varepsilon_t,$$

where  $\pi_t^e$  denotes expected inflation, while  $y_t$  denotes output in period  $t$  and  $y_e$  equilibrium output. The final term  $\varepsilon_t$  is interpreted as a “chock to inflation” (Jones 2014 p. 317). It reflects changes in the prices of oil or other inputs as well as changes in wages (p. 318), but according to Jones (p. 317) we should think of the term as being zero most of the time.

Moreover, if  $\pi_t^e = \pi_{t-1}$ , expectations are said to be *adaptive* and inflation is said to be *sticky*. Jones claims that the assumption of sticky inflation is a crucial one (pp. 309, 326), but the really crucial assumption is the approach to pricing. For this implies that a firm sets its prices based on its expectations of other firms’ prices, modified by its perception of the slackness of the economy, so that it will raise its prices somewhat less in a recession – “in an effort to recapture some demand” (p. 314) – while it will raise its prices somewhat more in a boom. And changes in costs affect pricing only through the price shock  $\varepsilon_t$ .

However, while the intuition of this price theory may be suggestive, I have seen no reference to microeconomic foundations. It also lacks empirical support. In fact, according to the survey on pricing in the U.S. in 1990-92 by Blinder et al., reported in detail in Blinder et al. (1998), expected inflation is irrelevant for individual pricing (Blinder et al. p. 98). In any case, a theory of pricing based on firms’ inflation expectations is very different from the theory of pricing developed in Chapter 1.

*Inflation, unemployment, and inflation expectations by wage setters*

In this case inflation is written as

$$(5) \quad \pi_t = \pi_t^E - \alpha(u_t - u_e),$$

---

<sup>27</sup> Jones (2014 p. 314).

where  $\pi_t$ ,  $\pi_t^E$ , and  $u_t$  refer to actual inflation rate, expected inflation rate, and unemployment rate for year  $t$ , while  $u_e$ , assumed to be constant in the medium run, captures all institutional and structural determinants of inflation (and also mark-ups and import prices).

While price setting in this case is assumed to follow mark-up pricing (Blanchard 2009 pp. 146-47), wage setting is assumed to depend on the price level expected by workers and firms (Blanchard p. 144). The intuition is as follows: “If workers expect the price level – the price of the goods they buy – to double, they will ask for a doubling of their nominal wage. If firms expect the price level – the price of the goods they sell – to double, they will be willing to double the nominal wage.” (Blanchard p. 145)

Now, with rational expectations,  $\pi_t^E = \pi_t$ , we always have  $u_t = u_e$ , while equation (5) with adaptive expectations,  $\pi_t^E = \pi_{t-1}$ , becomes

$$(6) \quad \pi_t - \pi_{t-1} = -\alpha(u_t - u_e).$$

In this case inflation increases if  $u_t < u_e$ , or decreases if  $u_t > u_e$ , while it is constant if  $u_t = u_e$ . This is why  $u_e$  often is called the non-accelerating (or non-increasing) inflation rate of unemployment, abbreviated NAIRU, or the equilibrium rate of unemployment. Thus, if unemployment is equal to its equilibrium rate, then inflation is stable (constant), both with rational and adaptive expectations of inflation.

Within this framework it is no wonder that a central bank targeting low and stable inflation wants to know what the equilibrium rate of unemployment is. For then it also concludes, if unemployment deviates from its equilibrium value, first, that inflation expectations are not rational and, second, how to change inflation, namely by an adjustment of the policy rate which changes aggregate demand and hence also output, employment and unemployment. However, if unemployment is at its equilibrium level, then (5) reduces to  $\pi_t = \pi_t^E$ , suggesting that inflation is determined by its expectations. And many central banks with an inflation target like 2% also attempt to establish inflation expectations equal to its target among wage setters, making wage formation and hence also price formation conditional on such expectations.

### *Inflation and price formation*

Now, as I have argued in this chapter, when firms are restricted by sales and not capacity, prices ( $p$ ) adjust to wages ( $w$ ) and other determinants according to

$$(7) \quad p = (1+m)(1+h)(w/a),$$

where  $a$  and  $m$  measure labour productivity and mark-up on variable costs, respectively, while  $h$  is a summary measure of the use of imports and other inputs (variable costs other than labour costs as a share of labour costs). Unless mark-ups or import costs change, inflation, as measured by a consumer-price index, is consequently due to wages increasing more than labour productivity in the consumer industry.

Prices of consumer goods, as well as inflation as measured by a consumer-price index, are determined by (7). A trade union for the consumer industry can consequently affect inflation as usually measured as long as production is restricted by sales at prices set by firms. And a rational union will not adjust its wage claims to its *expectations* of inflation but to its *target* for inflation, conditional on its predictions of labour productivity, mark-ups, and input prices – including prices of imports used as inputs in the consumer industry.

A central bank (CB) can reduce inflation as measured by a consumer price index by acting on effective demand through its policy rate if production is restricted by capacity in the consumer industry. In this case inflation can also be reduced without reducing employment. Otherwise, if production is restricted by sales, a CB can influence inflation only if it can affect wages in the consumer industry. But a CB can only control wage revisions indirectly, by creating and using unemployment as a disciplinary device when inflation is above its target.

Now, if “equilibrium unemployment” is estimated by a CB to be relatively high, then a CB should have very convincing empirical support for the assumption that wages are influenced by unemployment. For, if the effect of unemployment on wage revisions is negligible, then the only effect of increasing the policy rate will be to reduce production and increase unemployment without affecting inflation – as long as production in the consumer industry is restricted by sales and not capacity.

In any case it should be an important task for economic policy to ensure that wage formation is *not* affected by unemployment when production is restricted by sales in the consumer industry. In fact, as argued in Chapter 6, the crucial determinant of wage revisions is not unemployment but the interaction between employers and employees or their organizations within a framework of laws made by the parliament and regulations issued by the government.

## Chapter 4: Employment

In this chapter we shall see how employment depends on production and labour productivity, and how production is restricted by capacity or sales at prices set by firms. We shall also see how sales depend on spending (“effective demand”), and how spending depends on incomes generated by current production but also on savings and borrowings.

Thus, employment is determined in the product market by either capacity or sales at prices set by firms. Prices adjust to wages and import prices. Wages are normally revised at most once during a year, so we can take wages as given when prices, spending, production, and employment are determined.

Employment will depend on effective demand in an industry if production is restricted by sales at prices set by firms. Employment will depend on total effective demand if increases in total effective demand imply that effective demand increases in at least one industry where production is restricted by sales at prices set by firms. A crucial question is consequently to what extent production is restricted by sales and not capacity in an economy.

The determinants of a firm’s labour demand are summarized in Section 1, while aggregation of labour demand from a firm to its industry is discussed in Section 2. The relation between employment and effective demand is discussed briefly in Section 3 and more in detail, in a formal model within the framework developed in Chapters 1-2, in Section 4. Measuring the effect of matching problems on employment is the topic of Section 5. The evolution of employment in the long run is delineated in Section 6, while employment policy is discussed in Section 7. The last section relates the text to the literature.

### 4.1 Employment in a firm

Unless a firm’s output  $q$  and employment  $N$  are restricted by capacity  $k$  and labour productivity  $a$  (so that  $q = k$  and  $N = k/a$ ), employment is determined by

$$(1) \quad N = D(p)/a ,$$

where  $D(p)$  is the demand for the firm’s output at the price  $p$  set by the firm. And

$$(2) \quad p = (1+m)c = (1+m)(w+g)/a ,$$

where  $c$  is direct cost (variable cost per unit of output),  $w$  denotes the wage level (labour costs divided by employment) and  $g$  measures the costs of intermediate goods (variable costs other than labour costs divided by employment), while  $m$  is a mark-up on direct costs chosen by the

firm to cover indirect (fixed) costs and some profits, or to maximize profits or adjust to a given market price, as discussed in Chapter 1.

Note that labour productivity is a summary measure of the effect of labour on output at current technology and skills, including the use of real capital and intermediate goods. Note also that the cost of intermediate goods will reduce the sensitivity of labour demand to wage changes. In fact, the wage elasticity of labour demand is equal to the price elasticity of product demand multiplied by labour's share in total variable costs,  $w/(w + g)$ .

In practice there are, of course, some complications. Production does not adjust perfectly to sales unless sales precede production (production to orders). In markets where production precedes sales, as in most consumer markets, production will in general differ from sales, even if the change in inventories during a year may be negligible. And adjustment costs (costs of hiring and firing) will stabilize employment when sales are variable or hard to predict.

A firm's profits depend not only on variable costs but also on fixed costs. And the survival of a firm, and the evolution of its employment, depends not only on positive profits but also on a firm's ability to generate dividends for the firm's owners – and perhaps also bonuses for the firm's managers – as well as internal funds (retained profits) sufficiently large to absorb temporary losses and contribute to the financing of new investment.

#### **4.2 Employment in an industry**

Aggregation from a single firm to its industry is straightforward. Employment in the industry is determined as the sum of employment in its firms at a common market price. (In a market with differentiated goods prices may differ somewhat but the price level must be the same.) Mark-ups are not necessarily set independently by every firm. For example, in a market with price leadership all firms but one take the market price as given, and then the mark-up for a price taker is determined by the market price  $p$  set by the price leader and the direct cost  $c$  of the price taker,  $m = (p - c)/c$ , while the price leader sets that market price which maximizes its individual profits.

If production is restricted by capacity in every firm, then the industry's employment is also restricted by capacity and in fact determined by capacity and labour productivity (technology). And if production is restricted by sales in every firm, then the labour demand curve of the industry – relating the industry's employment to its wage level – has the same elasticity as the labour demand curve of a constituent firm if all firms are identical.

When firms are different the situation is more complicated, not only because the parameters in (2) may differ between firms, but also because production can be restricted by capacity in some firms and by sales in other firms. But in all cases there should be a negative relation between wages and employment, provided that production is restricted by sales for at least some firms, since then market prices will probably almost always depend on wages, implying that sales, production and employment will also depend on wages.

Focusing on wages, employment is always sensitive to wages if they are sufficiently high, so that production and employment are restricted by sales and not capacity. Then higher wages will by raising prices reduce sales, production and employment and the effect depends on the price elasticity of the demand for the industry's products but also on labour's share in total variable costs. However, the basic reason why there is a negative relation between wages and employment at the microeconomic level is that we can assume that the demand for an industry's products is independent of its firms' wages. And this assumption is not necessarily valid at the macroeconomic level.

The effect on aggregate employment of a general increase of wages depends on how higher wages add to product demand. And while higher wages will not increase demand for firms producing export goods, they may increase expenditures on consumer goods. But higher wages also imply higher prices on consumer goods, so the net effect on production and employment in the consumer industry is a priori ambiguous.

While a higher wage level does not *raise* employment in the export industry, it will not necessarily *reduce* employment – with two exceptions. Higher wages in exporting firms will raise prices and reduce sales, production and employment in firms which are price makers in global markets, provided their production is restricted by sales and not capacity. However, in exporting firms that are price takers in global markets, higher wages will reduce profits but not employment – unless wage increases are so large that some firms go bankrupt.

### **4.3 Employment and effective demand**

Aggregate employment depends on the relation between demand and capacity in every industry. In general an industry's regime – production restricted by sales or capacity – depends on the industry's product cycle: innovation, growth, maturity or decline. A new industry may find it possible to sell all it can produce at very high prices, while a declining industry may have excess capacity at the prices it sets. It can even happen that an industry's firms deliberately install or accept some excess capacity in order to be able to meet temporary

increases in demand without having to ration customers by either price or quantity (or delivery times).

If production is restricted by sales in an industry, employment is determined by sales and labour productivity, while sales are determined by spending at prices set by firms. While labour productivity depends on organisation, technology, real capital, and the use of imports as inputs in all industries, the determinants of spending differ substantially between industries. Thus, while wages are crucial for consumption, “animal spirits” may be decisive for investment and “international competitiveness” important for exports, while “deficit policy” may guide government spending.

There is an important distinction between exogenous and endogenous spending. *Endogenous* spending is financed by current income, that is, either income generated in current production (like current wages or current profits) or current transfers, including taxes and benefits. *Exogenous* spending is spending which is financed otherwise, either by dissaving (dependent on past income), or by borrowing (dependent on expectations of future income), or by other means, like taxes on property.

For example, in most firms investment is financed by past income (retained profits) or debt or equity and is consequently exogenous spending. Also, in most households consumption is financed out of current income and is consequently endogenous. The distinction is important, because without exogenous spending it may be difficult to increase production and employment. Since there is no production without sales and no sales without spending, there will be no production if spending depends entirely on income which does not exist until production has been realized. The distinction and interaction between exogenous and endogenous spending is explored in more detail in the next section, in a formal model within the framework developed in Chapters 1-2.

#### **4.4 Employment and effective demand in a formal model**

This section examines the relation between employment and effective demand in a more precise setting, that is, in a model with some precise – and restrictive – assumptions. In this model it will also be possible to clarify the distinction between exogenous (autonomous) variables and endogenous (non-autonomous) variables more in detail and show not only how exogenous variables determine endogenous variables “in equilibrium” but also what “equilibrium” means more precisely and how it can be attained.

Consider a closed economy (without imports or exports) and suppose that only households are taxed, that other transfer payments between sectors are negligible, that firms and

households don't borrow, and that the government's budget is balanced. With these assumptions we have for household consumption ( $C$ ) and government purchases from the market sector ( $G$ ), assuming that the wage level ( $w$ ) is the same in the market sector and the public sector,

$$(3) \quad C = wN - T_h - S_h,$$

$$(4) \quad G = T_h - wN_p,$$

where  $N$  denotes total employment in the economy,  $N = N_m + N_p$ , where  $N_m$  is employment in the market sector and  $N_p$  employment in the public sector, and where  $S_h$  denotes savings by households and  $T_h$  taxes paid by households. (Note that (3) also follows from (9) in Chapter 2 with  $F = D = A = B_h = 0$ .)

Suppose next that household taxes are proportional,

$$(5) \quad T_h = twN,$$

and that households' expenditures on consumption are proportional to current net income of employed workers,

$$(6) \quad C = \alpha(wN - T_h) = \alpha(1-t)wN,$$

where  $\alpha < 1$  if household savings are positive and  $\alpha > 1$  if they are negative. Hence

$$(7) \quad S_h = wN - T_h - C = (1-\alpha)(1-t)wN, \text{ and}$$

$$(8) \quad G = T_h - wN_p = twN - wN_p = twN_m - (1-t)wN_p = wN_m - (1-t)wN.$$

Moreover, total expenditures on output from the market sector are

$$(9) \quad E = C + G + I = wN - T_h - S_h + T_h - wN_p + I = wN_m - S_h + I,$$

where  $I$  denotes investment by firms in the market sector, while total output (GDP) in monetary terms is given by

$$(10) \quad Y = C + I + G + wN_p = wN + I - S_h,$$

which also can be written as

$$(11) \quad Y = wN + \Pi,$$

since aggregate profits  $\Pi$ , defined as firms' residual income,  $\Pi = E - wN_m$ , and with  $E$  according to (9), are given by

$$(12) \quad \Pi = E - wN_m = I - S_h.$$

Thus, even if aggregate profits depend negatively on wages, they are *ultimately* determined by investment and household savings. Note that (12) is a special case of the profit equation derived in Chapter 2.

Of course, (12) can also be written as

$$(13) \quad I = S_h + \Pi ,$$

implying that investment is equal to savings, where  $S_h$  is saving by households and  $\Pi$  is saving by firms (since there are no dividends in this simple model). But this means that investment is *financed* by savings, not that it is *determined* (restricted) by savings. This distinction reflects two different decisions by firms: a primary (exogenous) decision on investment and a secondary (endogenous) decision on its financing. Note also that

$(S_m - B_m) + S_h = 0$  according to the fundamental identity in Chapter 2, so (13) implies (implicitly) that  $I = \Pi + B_m - S_m$ , showing how investment in practice can be financed by a combination of borrowing and dissaving out of accumulated profits, even if current profits will replenish firms' money holdings.

Next we note that, according to (10),  $Y = wN$  if  $I = S_h$ , and in general that

$$(14) \quad Y = wN + I - S_h = wN + I - (1 - \alpha)(1 - t)wN = wN(1 + \gamma), \text{ where}$$

$$(15) \quad \gamma = \beta - (1 - \alpha)(1 - t), \text{ with}$$

$$(16) \quad \beta = I/wN .$$

It follows that we can write consumption, government purchases, and investment as a share of GDP as

$$(17) \quad C/Y = \frac{\alpha(1-t)}{(1+\gamma)} ,$$

$$(18) \quad G/Y = \frac{wN_m - (1-t)wN}{wN(1+\gamma)} = \frac{(N_m/N) - (1-t)}{1+\gamma} ,$$

$$(19) \quad I/Y = \beta/(1+\gamma) .$$

Note also that wages and profits as shares of total income are given by

$$(20) \quad \frac{wN}{Y} = \frac{1}{1+\gamma} ,$$

$$(21) \quad \frac{\Pi}{Y} = \frac{\gamma}{1+\gamma} .$$

Thus, the model determines the *distribution* of output or income between important categories, and in fact not only the distribution of *nominal* output and income, but also (indirectly and approximately) the distribution of *real* output or income, even if this distribution also depends on how real output is measured in practice in national accounts.

Moreover, this is a model of full employment if employment in the market sector corresponds to full capacity utilization. In this case employment in the market sector is determined by capacity and labour productivity, while prices are determined by market clearing, and spending only affects inflation.

However, if production is restricted sales at by prices set by firms, then prices in the market sector, when variable costs are independent of imports, can be written as

$$(22) \quad p = (1+m)c = (1+m)w/a ,$$

where  $m$  is a mark-up and  $a$  is labour productivity,  $a = q/N_m$ , where  $q$  is output. In this case total purchases from the market sector can be written as

$$(23) \quad pq = (1+m)(w/a)aN_m = (1+m)wN_m ,$$

and since  $pq = E$  we obtain from (23) and (9):

$$(24) \quad (1+m)wN_m = wN_m - S_h + I .$$

It follows from (24) and (12) that

$$(25) \quad m = (I - S_h)/wN_m = \Pi/wN_m .$$

Of course, since profits are generated by mark-ups on wage costs, the (average) mark-up must be related to (aggregate) profits, even if no firm can know the relation between aggregate investment and the wage bill beforehand. Note, however, that (25) does not imply that aggregate profits are determined by mark-ups, but that the *average* mark-up is determined by aggregate profits – which are determined by firms' investment and households' savings (ignoring export surpluses and government deficits in this simple model).

Moreover, employment in the market sector, when production is restricted by sales at prices set by firms, is determined as the sum of employment in firms producing consumer goods ( $N_m^c$ ), investment goods ( $N_m^i$ ), and goods purchased by the public sector ( $N_m^p$ ),

$$(26) \quad N_m = N_m^c + N_m^i + N_m^p .$$

Now, according to (23), (6), and (8), employment in firms producing consumer goods, investment goods, and public goods are determined by the following equations:

$$(27) \quad \alpha(1-t)w(N_m^c + N_p) = (1+m_c)wN_m^c ,$$

$$(28) \quad I = (1 + m_i) w N_m^i ,$$

$$(29) \quad t w N_m - (1 - t) w N_p = (1 + m_p) w N_m^p ,$$

where  $m_c$ ,  $m_i$  and  $m_p$  are mark-ups in the three sectors.

Thus, market employment is determined by a system of equations. Investment needs planning for its implementation and this means that employment in the production of investment goods should be established early in a year, so that  $N_m^i$  is determined by (28) at an early stage. Hence  $N_m^i$  can be taken as given in both (27) and (29), but these are two equations in two unknowns,  $N_m^c$  and  $N_m^p$ . (Recall that while  $N_p$  – employment in the public sector – can be taken as exogenously given by political decisions,  $N_m^p$  is employment in the market sector involved in the production of goods purchased by the public sector).

How will employment adjust to “equilibrium” defined by equations (27) - (29)? One possibility is that the government adjusts its purchases from the market sector and the concomitant  $N_m^p$  until equilibrium is attained. This can be done by predicting the equilibrium value of  $N_m$  and adjust government purchases and hence  $N_m^p$  until the budget is balanced for this value of market employment, so that eq. (29) is satisfied. And summing equations (27) - (29), the government can obtain an equation for  $N_m$ , namely

$$(30) \quad \alpha(1-t)(N_m + N_p) + I/w + tN_m + (1-t)N_p = (1+m)N_m ,$$

assuming for simplicity that  $m_c = m_i = m_p = m$ . In this case we have

$$(31) \quad I/w + (\alpha + 1)(1-t)N_p = (1 + m - \alpha(1-t) - t)N_m , \text{ so that}$$

$$(32) \quad N_m = \frac{I/w + (\alpha + 1)(1-t)N_p}{1 + m - \alpha(1-t) - t} ,$$

or, substituting  $m = ap/w - 1$  according to (22),

$$(33) \quad N_m = \frac{I/p + (\alpha + 1)(1-t)(wN_p/p)}{a - (\alpha + (1-\alpha)t)(w/p)} ,$$

which shows how market employment “in equilibrium” depends on the real value of investment as measured by  $I/p$ , but also on the real value of the wage bill in the public sector,  $wN_p/p$ . Note that the “multiplier” depends on the difference between labour productivity ( $a$ ) and the real wage ( $w/p$ ) and is exactly equal to  $1/(a - w/p)$  if  $\alpha = 1$ .

This simple model shows how employment depends on effective demand (spending). It also shows how important the distinction between exogenous and endogenous components of effective demand is. More precisely, it shows how dependence of spending on current employment introduces a “feed-back loop” which magnifies the effect of variations in exogenous components on employment.

The model also suggests that aggregate employment can depend on the real wage ( $w/p$ ) even if employment at the firm level does not (as shown in Section 1). In this simple model aggregate employment will even depend positively on real wages, all else being equal, including investment. This may appear strange, assuming that investment depends on profits and that profits depend negatively on wages for a firm or an industry. However, as we have seen in Chapter 2, aggregate profits do not ultimately depend on wages.

#### **4.5 The effect of matching problems on employment<sup>28</sup>**

Some hires are made more or less instantaneously by firms, for example by recalling workers previously laid off or by offering jobs to spontaneous job applicants. In other cases there is no existing pool of job applicants which a firm can turn to. Instead the firm has to attract job applicants by advertising its demand for personnel in newspapers or other media, by placing job orders with a public or private employment agency, or by contacting potential candidates directly. And then vacancies understood as *recruitment processes* arise.

Moreover, firms start recruiting in anticipation of future needs. If, for instance, a separation can be anticipated and a replacement made before the separation, then replacement is instantaneous even if recruitment is not. But otherwise vacancies understood as *unfilled jobs* (unmet labour demand) exist from the day the employer wants the worker to start to the day the worker starts. An unfilled job can consequently be interpreted as an *unplanned dip in employment* or an “empty chair”.

Thus, firms create “vacancies” in one sense (recruitment processes) in order to avoid “vacancies” in another sense (unfilled jobs). Because of this ambiguity, both in the economics literature and in everyday language, I will refer to vacancies as recruitment processes as *job openings* (or openings) and vacancies as unmet labour demand as *unfilled jobs*. The distinction is important, since it is only unfilled jobs which reduce employment by making the number of employees less than the number of jobs.

---

<sup>28</sup> This section is based on Farm (2017c).

In the Swedish business survey on vacancies unfilled jobs are defined more precisely as *unoccupied job openings which are available immediately*. This is a definition which excludes occupied job openings and job openings to be filled later, in the same way as the definition of unemployed workers in labour force surveys excludes job seekers with a job and job seekers without a job who cannot start work until later.

Moreover, in the Swedish survey unmet labour demand (unfilled jobs) is measured indirectly, as unmet labour supply (unemployed workers) is measured in labour force surveys, by a succession of questions. More precisely, unfilled jobs are defined as a subset of job openings obtained by eliminating first “occupied job openings” and then “unoccupied job openings which are unoccupied because no work is wanted by the employer until later”. And “occupied job openings” exist when, during recruitment of new workers, the corresponding jobs are occupied by retiring workers or substitutes until replacements or permanent personnel have been hired.

Now, according to the Swedish vacancy survey, which measures not only job openings but also unfilled jobs, only about 40 per cent of job openings are also unfilled jobs. On the other hand, since 2000 the rate of unfilled jobs has averaged 0.5 per cent of employment in the private sector in Sweden. Thus, merely the time it takes to recruit workers reduces employment by creating a gap between jobs and employment of about 0.5 per cent of employment on average. And the rate of unfilled jobs depends on the phase of the business cycle, varying from 1.1 per cent in the last two quarters of 2000 – and the second quarter in 2017 – to 0.3 per cent during the recession in 2009, according to Statistics Sweden.<sup>29</sup>

#### **4.6 Employment in the long run**

Labour demand in the long run – or more precisely the development of employment over time – is determined by the history of its growth, that is, by the accumulation over time of yearly changes. Such changes are based upon changes in market conditions, including changes in effective demand, capacity, technology, wages and other variable costs, and mediated through adjustment costs and expectations. More precisely, when the determinants of the yearly demand for labour change, as specified in Section 1, employment will change, but not necessarily immediately, because of adjustment costs and uncertainty about how permanent perceived changes are.

---

<sup>29</sup> See “Konjunkturstatistik över vakanser” at [www.scb.se](http://www.scb.se).

A basic question is how employment responds to technological progress, including in particular increasing labour productivity. According to eq. (2) higher labour productivity will reduce the prices set by a firm and its industry and hence increase production, other things equal, as long as production is restricted by sales. The net effect on employment will depend on the elasticity of product demand,  $\eta = -(dD/D)/(dp/p)$ , since  $dp/p = -da/a$  according to (2) and

$$(34) \quad dN/N = dD/D - da/a = -\eta dp/p - da/a = (\eta - 1) da/a .$$

Thus, employment will increase with increasing labour productivity if product demand is sufficiently elastic ( $\eta > 1$ ). If increasing labour productivity is accompanied by increasing effective demand, firms may find it profitable to increase employment even if product demand is inelastic ( $\eta < 1$ ). Thus, increasing labour productivity will reduce employment with certainty only if production is restricted by capacity and not sales, since in this case  $N = k/a$ .

An industry can also increase its employment by increasing its capacity – if the industry is producing at full capacity. Changes of capacity will, of course, depend on the phase of the industry's product cycle (innovation, growth, maturity or decline), but also on whether the industry is producing at full capacity or not. "Animal spirits" can be decisive during an industry's innovation phase. Strategic considerations will always enter the picture, assuming that an industry's firms cannot coordinate their investment decisions. Particularly important may be the existence or entry of low-cost producers with large capacities, aiming at large sales due to elastic product demand at low prices. On the other hand, in an industry where product demand is expected to decline in the long run, even firms producing at full capacity may be tempted to reduce capacity.

The growth of market employment has been facilitated by labour reserves generated by higher labour productivity in farming and homework due to capital goods produced by the market sector. At the same time employment in the public sector has also grown, so that employment in the formal economy (comprising the market sector and the public sector) in many countries has reached 75 - 85 per cent of the population in working age. But what happens when employment in the market sector no longer can grow at the expense of work in traditional agriculture or homework?

Of course, some activities in the public sector can be outsourced to the market sector through public procurement, but there are limits to such transfers. The growth of employment in some new industries can be interpreted as the outsourcing of some functions from traditional industries, implying merely the transfer to specialized firms of personnel doing the

same as before, that is, a more advanced division of labour. Otherwise an industry can expand its employment only at the expense of employment in another industry, and labour mobility between industries usually involves costly retraining and costly moving to new dwellings.

Increasing difficulties to expand sales in an industry may stimulate investment in labour-saving technologies. But without increasing market employment the growth of effective demand based on growing wage bills in the market sector may stagnate. And effective demand may stagnate in a mature market economy even for other reasons.

First, effective demand for investment goods may stagnate or decline unless a new industrial revolution, like the introduction of information and communication technology (ICT), makes the profitability of investment in real capital as obvious as before. Otherwise firms may choose to invest abroad in real capital or they may choose to invest in financial capital. Recall that investment also has the fundamental function of validating expectations of profits, as emphasized in Chapter 2, implying that a decline in investment may reduce profits so much that some firms leave the market unless effective demand is sustained by excess borrowings by households or the government – or by exports exceeding imports. Second, effective demand for consumer goods may stagnate because of increasing housing costs or increased savings or increased inequality, reducing effective demand from households with stagnating incomes for obvious reasons and reducing effective demand from rich households because of increased savings. Third, government spending on employment in the public sector or on goods and services from the market sector, including investment in infrastructure, may stagnate because of increasing budget deficits and public debts.

#### **4.7 Employment policy**

Even if market employment so far has been steadily increasing in most countries, dips in its evolution have also occurred. A fall in *total* market employment may be the net outcome of on one hand industries with declining employment and on the other hand industries with rising employment, reflecting a structural transformation of the market sector. For example, investment in real capital is often unstable and much of its variability can be explained by the decline of some industries and the rise of others as the economy develops.

If employment in some industries are declining while employment in other industries are expanding, then a government can facilitate the transfer of employment from old to modern industries by labour market policy and housing policy. For an industry with permanently declining employment other policy measures may also be contemplated, depending on the causes of decline. If the fall in employment is due to increasing labour productivity, a

government does not normally intervene. If the fall in employment is due to output being replaced by imports, such a substitution can be prevented by tariffs on imports, unless such measures are excluded by international agreements on tariffs and trade. A decline in employment can also be moderated by various forms of subsidies, including income support, particularly in agriculture, where it can be motivated by a wish to preserve the ability to produce food in an international crisis.

Suppose next that the decline of employment in an industry is due to falling demand for its products. If the industry is producing consumer goods, the decline may only be a reflection of changing preferences. If it is, economic policy is normally restricted to labour market policy facilitating the transfer of employment to expanding industries. But if the decline applies to most industries producing consumer goods the situation is more complicated. If the decline depends on stagnating wages, effective demand for consumer goods can be supported by lower taxes. But if the decline depends on increased household savings (or repayment of debt), then lower taxes may only add to these withdrawals of effective demand unless they are targeted towards low-income households. And if all households want to increase their financial security by additional savings, it could be argued that the government should accept this and attempt to stimulate the transfer of some employment from producing consumer goods to other activities, including the maintenance or construction of vital infrastructure.

On the other hand, some households may want to spend some of their surplus money on improving education for their children, or on health care or elderly care, that is, on services which so far have been public services in many countries. This would create a demand for privatisation of some public services and consequently constitute another opening for market employment. Thus, even if the government historically has initiated individual services like education, it is possible that the development of some of these services at some stage can be handled by the market. At the same time there can be important new activities which the market mechanism cannot initiate today and which consequently should be developed within the public sector, particularly if employment in the market sector stagnates. And in this case, with stagnating market employment, perhaps a government should be not only the “lender of last resort” (to support liquidity) or the “borrower of last resort” (to support effective demand) but also the “employer of last resort”, as argued, for example, by Wray (2012).

Economic policy can also support investment in the market sector. In fact, governments have often tried to counteract a decline in investment by stimulating investment in various ways, including subsidies of inventories or capacity or technology. Or they have tried to

compensate a decline in private investment by a temporary increase in public investment, particularly in infrastructure.

On the other hand, economic policy can also support investment in basic industries, particularly in emerging economies, both directly and indirectly. Perhaps the most important indirect support is to allow banks to supply loans by creating money. A government can also support investments in the market sector indirectly by supporting their profitability, either by supporting exports or facilitating household debt or increasing public debt, as discussed in Chapter 2.

#### **4.8 Relation to the literature**

In traditional microeconomics the labour demand of a firm does not depend on the demand for the firm's products, only on the production function and the real wage  $w/p$ , where  $w$  denotes the firm's wage level and  $p$  the price of its output. And this result follows from the idea that a profit-maximizing firm increases employment until the value of the marginal product of labour is equal to the wage rate.

On the other hand, labour demand is a derived demand. The dependence of labour demand on product demand through a price related to costs is also emphasized in most textbooks *at the industry level*. But if labour demand is a derived demand at the industry level, it should also be a derived demand at the firm level. And this is indeed the case – in markets where prices are set by firms and not by a market-clearing auctioneer or process.

The intuition is as follows. In practice a firm does not choose employment but the price of its output before trade can start. And then it adjusts its production to its sales and its employment to its production. Thus, a firm's employment must depend on the demand for its products at the prices it sets. This is explained in detail in Farm (2017b), but it also follows immediately from eq. (1) in this chapter.

#### *Equilibrium employment in classical models*

In traditional macroeconomics, employment is determined simultaneously as wages, at least “in equilibrium” and “in the medium run”. Classical models of the labour market start from the idea that price and quantity in all markets are determined by equality between supply and demand. Thus, in classical models the wage level and the employment level are determined simultaneously by the intersection between a downward-sloping labour demand curve and an upward-sloping labour supply curve, where labour demand and labour supply are assumed to

be functions of the “real wage”, that is the nominal wage level divided by a price index. Thus, in classical models employment and real wages are determined by supply and demand in the labour market, while output is determined by employment and labour productivity.

Of course, aggregate employment and the real wage level are not determined by equality between supply and demand on every day, as price and quantity are determined in financial markets, but only “in the medium run” or “on average”. More precisely, it is assumed that excess supply – that is, unemployment – implies a decreasing real wage, while excess demand (overemployment) implies an increasing real wage, so that any deviation from market clearing will be corrected by “market forces”, implying a return to market clearing “in the medium run” after a single disturbance, and market clearing “on average” when disturbances are frequent (and stochastic). Thus, in this framework there is no unemployment in equilibrium, at least not if wages are “flexible”.

In practice it is assumed that the real wage adjusts only if the unemployment rate deviates from a certain positive value, often called the *natural rate of unemployment*. And it is implicitly assumed that real wages adjust by adjustment of nominal wages to the prices of consumer goods, as in the nineteenth century in England, where nominal wages were adjusted to the price of imported grain.

Now, this adjustment process does suggest an employment concept different from actual employment, namely “equilibrium employment”, towards which employment tends after a single disturbance, or around which employment fluctuates when disturbances are frequent (and stochastic). And while employment “in the short run” – out of equilibrium – may be determined in the product market, employment “in the intermediate run”, that is, “equilibrium employment” is determined in the labour market by equality between supply and demand.

However, if employment is determined in the product market “in the short run”, this means that employment is determined in the product market not only in the current year but in every year. And if employment is determined in the product market during every year, its average over the near future is also determined in the product market. What we have is simply two different and conflicting theories, and they cannot both be true at the same time, not even when relating them to different time periods. And while market clearing is applicable to markets for commodities or securities, it is simply not relevant in labour markets.

#### *Equilibrium employment in modern models*

However, in modern monographs and textbooks, including, for example, Layard et al. (1991), Blanchard (2009), and Carlin & Soskice (2015), the labour-demand curve is replaced by a

*price-setting curve*, and the labour-supply curve is replaced by a *wage-setting curve*. The intersection of these curves, where both price setters and wage setters are content with the prevailing real wage, is supposed to determine employment “in equilibrium” or “in the medium run”, as distinct from actual employment “in the short run”.

The price-setting curve introduces mark-up pricing in macroeconomics, implying that prices adjust to nominal wages, not the other way around, as in classical models. It follows from the insight that the traditional model of labour demand for a non-competitive firm must be interpreted as a price equation.<sup>30</sup> In fact, according to eq. (17) in Chapter 1, and if firms are restricted by sales and not capacity, prices ( $p$ ) adjust to wages ( $w$ ) according to

$$(35) \quad p = (1+m)(1+h)(w/a),$$

where  $a$ ,  $h$ , and  $m$  measure labour productivity, variable costs other than labour costs as a share of labour costs, and the mark-up on variable costs, respectively.

Aggregation over firms will weaken this relationship, but according to (35) the real price level  $P/W$  (where  $P$  measures the price level and  $W$  the wage level) is uncorrelated with employment, suggesting a horizontal price-setting curve (with price independent of employment),

$$(36) \quad P/W = (1+m)(1+h)/a,$$

where  $h$  measures the use of import goods in the production of final goods. Thus, the price-setting-wage-setting model, or PSWS-model for short, essentially offers a *norm* for wage formation, with (36) defining a realistic level of real wages in terms of labour productivity, input structure and mark-ups in the consumer industry,

$$(37) \quad W/P = \frac{a}{(1+m)(1+h)}.$$

Now, the complete PSWS-model also assumes that some unemployment is necessary to make wage setters accept the real wage level determined by current conditions in the consumer industry. It has even been argued that if prices on imports like oil rise so much that the feasible real wage falls, then: “Getting workers to accept the lower real wage requires an increase in unemployment” (Blanchard 2009 p. 175). Or, as argued by Layard et al. (1991 p. 12), if unemployment is “too low” it will start a wage-price spiral when workers try to obtain a real wage which is not feasible. Inflation is stable only if unemployment is “sufficiently high” or “in equilibrium”.

---

<sup>30</sup> But this insight should also lead to an updated model of labour demand, as in Farm (2017b).

Thus, “in the medium run”, employment is determined in the labour market even in PSWS-models. More precisely, employment is determined “in equilibrium” by the labour force and the “equilibrium rate of unemployment”, while “equilibrium output” is determined by “equilibrium employment” and labour productivity.

However, as already noted, employment cannot be determined each year by two different theories, and if employment is determined each year in the product market, its average or final value during some years constituting “the medium run” must also be determined in the product market.

Moreover, realizing that prices are adjusted to wages according to (36) in industries producing and distributing consumer goods, rational workers or trade unions should realize that what is important is not *expectations* of the price level but *information* on the determinants of inflation. In fact, not only prices of consumer goods but also rational expectations of inflation are determined by (36). Thus, a trade union in the consumer industry should not demand compensation for the *expected* rate of inflation, but, as elaborated in Chapter 6, adjust its wage claims to its predictions of labour productivity and import prices in the near future, and to its *desired* rate of inflation.

## Chapter 5: Unemployment

Unemployed people are usually defined as persons without employment looking for jobs. Not all people without employment are looking for work and they may not even be able or willing to undertake employment. Ability may be restricted by illness or disabilities or other activities, like education or homework. Willingness may depend on other sources of income. In fact, defining and measuring “unemployed” as distinct from other people (in working age) without employment is an intricate problem. It cannot always be identified with “looking for work”, as discussed in Section 1.

Section 2 relates the stock of unemployment to its inflows and outflows, while the relation between unemployment and wages is discussed in Section 3. The idea that a certain level of unemployment is necessary to prevent unacceptable inflation is then examined in Section 4, while the trade-off between unemployment and inflation is discussed in Section 5. What policy can do to reduce unemployment due to a fall in effective demand is discussed in Section 6, while Section 7 focuses on unemployment due to other events. The last section relates the text to the literature.

### **5.1 Defining and measuring unemployment**

In labour force surveys unemployed persons are usually defined as people without employment who are “looking for jobs”, and they are measured by a question on job search during the last month. But “discouraged workers”, who no longer find job search meaningful, are also registered in labour force surveys, suggesting a broader definition of unemployment.

Of course, not all people without employment are “unemployed”, particularly not if they are “too young” (in school) or “too old” (retired). Some persons in working age may be students at universities or colleges. Some persons in working age who are not students may not be able or willing to undertake employment. Ability may be restricted by illness or disabilities. And willingness may depend on other sources of income or other possibilities to earn a living, for example in traditional forms of agriculture or in homework. In emerging economies the possibility for workers to return to subsistence farming when market employment falls makes it particularly difficult to define and measure unemployment. In advanced economies the option of subsistence agriculture is no longer available and the distinction between unemployed and other non-employed is more important.

An individual's "ability to work" may be restricted by other things than health, either by firms' requirements or the individual's requirements. Depending on education and job experience, an individual is qualified for employment only in some jobs in some firms. On the other hand, an individual's labour supply may be restricted to some firms in the local labour market for several reasons, including a partner's current employment. And without an overlap between the set of firms which find an individual properly qualified and the set of firms acceptable to an individual, an individual may sooner or later stop looking for work.

In general, six characteristics of a non-employed person (in working age) appear to be important, namely education, job experience, financial support, job search, and activity apart from job search. Health is a matter of degree and inability to work may be temporary or permanent and need not apply to all types of work. Non-employed persons can be financially supported, apart from wealth, by pensions, student loans, unemployment benefits or social assistance, or otherwise by family and friends or charitable institutions. Apart from job search of various kinds, non-employed persons who are not students may spend their time on homework, including time-consuming care of children or elderly which has to be replaced by other forms of care if a job offer is obtained and accepted.

Aggregate employment will decrease when employees retire or leave employment for other reasons – unless employers recruit replacements from the pool of non-employed. In this context a crucial question seems to be how the probability of transfer from non-employment to employment depends on the characteristics of the non-employed. Are personnel recruited entirely from the pool of job searchers or are some former employees simply recalled to work or contacted directly by employers? How often are students recruited directly from school? And how often are non-employed persons with disabilities recruited?

Moreover, the possibility of transfer from non-employment to employment depends on aspects of both demand and supply. First, how many non-employed persons are qualified for employment even if they are not presently looking for work? An estimate of the number of such persons presupposes information on education, job experience and health status for all non-employed. Second, how many "non-employed qualified for work" are also willing to accept a job offer, apart from job searchers? An estimate of the number of such persons presupposes a question in labour force surveys on the willingness to accept a job offer if offered a job. As emphasized, for example, by Blanchard (2009 p. 138), the unemployment rate as defined in traditional labour force surveys is not necessarily the best estimate of the number of people available for work.

Information on “non-employed who are both qualified for employment and willing to accept a job offer” would be important information on the “labour reserve” and may even suggest an alternative definition of unemployment. In this context information on how often an employer’s job offer is rejected by a non-employed person would give basic facts on “willingness to accept a job offer”, but I have found no systematic information on this issue in the literature.

Even if employment is important for most people, and not only because of its monetary rewards, some people may prefer and choose other alternatives in a rich society. And I am not thinking of aversion to employment among *rentiers*. Nor am I thinking of unemployed workers rejecting job offers at the same time as they are collecting unemployment benefits. Instead I am thinking of people who deliberately choose more “leisure” than normal employment allows.

Sometimes this may also be leisure as usually defined, for example when a person chooses early retirement. Sometimes people choose part-time employment because they have to, or prefer to, take care of their children or relatives. But people may also choose a deviant lifestyle, including subsistence farming or home production with a minimum of goods or services purchased in the market. In a rich society such a lifestyle can also be financed by some temporary or seasonal employment. And even if such labour supply on one hand can be matched by temporary or seasonal labour demand, it will, on the other hand, reduce total labour supply from non-employed people. It remains to see if such a reduction of total labour supply will be an important restriction on the growth of market employment in the future – or an important complement to stagnant market employment.

## 5.2 Unemployment and its flows

The unemployment rate is usually defined as the number of unemployed persons at a particular point in time divided by the number of people in the “labour force”, defined as the set of employed or unemployed persons.<sup>31</sup>

It is tempting to think of the stock of unemployment at a particular point in time as *determined* by past inflows and outflows, at least “in the long run”, or “in equilibrium”, or “on average”. It is also obvious that for an *individual* the causes of unemployment can fruitfully be divided into, first, causes of *becoming* unemployed and, second, causes of *remaining*

---

<sup>31</sup> In practice the unemployment rate for a month may refer to the end of the month or it may be an average over the month. In labour force surveys the reference period may be a week instead of a day and unemployed persons may be defined as persons looking for work who have been non-employed the whole week.

unemployed. But what is true for an individual is not necessarily true for a group of individuals.

In fact, even if policy measures can reduce the risks of becoming or remaining unemployed for certain groups of people, they can only reduce total unemployment by raising total employment or reducing the number of non-employed looking for jobs. The second approach has sometimes been implemented by facilitating early retirement, but otherwise unemployment can only be reduced by increasing employment.

Of course, employment can be increased by labour market policy measures for unemployed people, like retraining, employment subsidies, or intensified job search – even without increasing effective demand at the same time. However, in order to show effects on aggregate employment one has to show that these measures will raise employment *either directly*, by reducing matching problems so much that the number of unfilled jobs also is reduced, as elaborated in Chapter 4, *or indirectly*, by reducing product prices for some firms so much that sales and hence also production and employment are increased.

Product prices can be reduced by reducing costs and costs can be reduced by reducing labour costs, including recruitment costs. Employment subsidies will also reduce labour costs, while retraining of unemployed can reduce recruitment costs but also the rate of unfilled jobs if properly designed. Increasing the search intensity of the unemployed may also reduce the rate of unfilled jobs, particularly in a boom, when the rate of unfilled jobs may be large.

However, reducing aggregate unemployment may not be the only motive of labour market policy. Affecting the *distribution* of unemployment may also be important, in particular the distribution between short-term and long-term unemployment. Short spells of unemployment between jobs are not as serious as long spells, particularly not if long spells end by unemployed persons in working age leaving the labour force.

Of course, in some labour markets unemployment spells may be long simply because job offers are rare and qualified workers few. And then an unemployed person knows that a job offer matching her qualifications will sooner or later pop up. But in a labour market with a constant flow of job offers and an even larger flow of job applicants, a long unemployment spell may indicate less qualification than other applicants have to an employer, who then routinely may exclude long-term unemployed even from interviews.

Some policy measures may increase the chances for long-term unemployed to obtain not only an interview with an employer but also a job offer, even in labour markets with many applicants for every job vacancy. And even if such assistance does not increase aggregate employment, it may reduce long-term unemployment and hence also the “seriousness” of

unemployment as measured, for example, by the share of unemployed with unemployment spells longer than six months.

Note also that even if reducing the generosity of the unemployment-insurance system may increase the search intensity of the unemployed and also reduce the duration of unemployment for some groups, aggregate unemployment will only be reduced if the rate of unfilled jobs is reduced – or, perhaps, if lower unemployment benefits increase employment by reducing wages.

### **5.3 Unemployment and wages**

Does unemployment exist because wages are too high? Or, more precisely, will lower wages reduce unemployment by increasing employment?

Suppose that lower wages in an industry does not reduce the demand for the industry's products. Then lower wages will increase sales, production and employment in the industry if – and only if – lower wages also reduce product prices. The effect is substantial, however, only if demand for the industry's products is elastic and labour's share in total variable costs is large, according to Chapter 4.

This argument also applies to the export industry, but only for firms that are price makers in global markets (since firms which are price takers will only face higher profits). It does not necessarily apply to the consumer industry, since lower wages will also reduce the demand for this industry's output. In fact, even if lower wages in an industry does not affect the demand for the industry's output, it will always reduce the demand for consumer goods and hence also employment in the consumer industry.

Thus, lower wages in an industry have both a direct effect on employment (by reducing prices and increasing the industry's sales, production and employment) and an indirect effect (by reducing the effective demand for consumer goods and hence employment in the consumer industry). The net effect is ambiguous and ultimately an empirical question.

### **5.4 Unemployment and inflation**

Unemployment should, of course, be low, but a common qualification is that it should not be so low that it threatens to raise inflation above a certain target, for example 2 per cent per year. This restriction is based on the assumption – examined in detail in Chapter 6 – that wage formation, and hence also price formation, is sensitive to unemployment. For only then will an unemployment rate below a certain critical level – which I will call the *necessary rate of unemployment* (NERU) – raise inflation above its target.

This assumption is related to the presumption that there is a “natural rate of unemployment” such that any deviation from it will soon be corrected by adjustment of the wage level. More precisely, unemployment below the natural rate is assumed, firstly, to increase wages, while unemployment above the natural rate is assumed to decrease wages. But it is also assumed, secondly, that an increasing wage level will reduce employment and hence raise unemployment, while a decreasing wage level will raise employment and hence reduce unemployment.

In practice, however, the effect of unemployment on wages may be rather weak, as argued more in detail in Chapter 6, while the effect of the wage level on aggregate employment is ambiguous, as argued above. Moreover, even if the natural rate of unemployment exists, the rate of adjustment, both the adjustment of wages to unemployment and the adjustment of employment to wages, may be too slow for deviations from the natural rate to have any effect on unemployment in practice, at least compared to all other determinants of unemployment.

However, a “necessary” rate of unemployment may exist even if there is no “natural” rate of unemployment. Of course, this presupposes that wage formation is sensitive to unemployment. But it does not presuppose that deviations of wage inflation from its necessary level – corresponding to the inflation target – *automatically* triggers adjustment of unemployment towards its necessary level. Instead it presupposes that deviations from the necessary rate of unemployment triggers adjustment of unemployment by the central bank, which by adjusting its policy rate can adjust effective demand and hence also employment and unemployment. In this case the “necessary” rate of unemployment will also be an “equilibrium” rate of unemployment in the sense that any deviation from it will trigger corrections by the central bank.

Now, if this “equilibrium rate of unemployment” is estimated to be relatively high, then the central bank should have very convincing support for the assumption that wages are influenced by unemployment. For, if the effect of unemployment on wage revisions is negligible, then the only effect of increasing the policy rate will be to reduce effective demand and increase unemployment without affecting inflation – as long as production in the consumer industry is restricted by sales and not capacity.

### **5.5 On the trade-off between unemployment and inflation**

In any case, the basic problem for a central bank is not the relation between wage formation and unemployment but the relation between its policy rate on one hand and inflation and unemployment on the other hand. Is there always a trade-off between unemployment and

inflation, or is it possible to reduce unemployment by reducing the policy rate without at the same time increasing inflation? And more generally: can economic policy by increasing effective demand reduce unemployment without increasing inflation?

First, if production is restricted by capacity at prices set by firms in the consumer industry, then increasing effective demand would only increase inflation without reducing unemployment. Second, if production is restricted by sales, then inflation depends on the evolution of labour costs, costs of other variable inputs, mark-ups, and labour productivity in the consumer industry, according to eq. (1) in Chapter 3. It follows that a target for price inflation implies a target for wage inflation, conditional on changes in labour productivity, input prices and mark-ups. In other words, a target for a change of the wage level must be adjusted to the target for inflation and anticipated changes in labour productivity, input prices and mark-ups. And in the simplest case, with no trend in the development of input prices and mark-ups, the target for wage increases cannot exceed the target for price increases by more than the predicted increase in labour productivity in the consumer industry.

Thus, a target for inflation (as measured by a consumer price index) always implies a corresponding target for an increase of the wage level when wages are revised in the consumer industry. I assume that a central bank (CB) can estimate this target and also observe actual changes of the wage level. If wage inflation is higher than its target, a CB can raise its policy rate if it believes that a higher policy rate will not only increase unemployment but also reduce wage claims by employees or their trade unions and wage concessions by firms or their organisations.

Alternatively, if all partners in the process of revising wages respect not only the target for price inflation but also the corresponding target for wage inflation, a CB can focus entirely on the effect of its policy rate on unemployment. Of course, this presupposes coordination in wage setting and probably also negotiations between trade unions and employer organisations that have a decisive influence on the change of the wage level in the consumer industry. But such institutions do exist in many countries.

In any case, monetary policy should depend on how wages are revised. If wage setting is coordinated in such a way that the outcome is consistent with the inflation target, then a CB can focus entirely on *unemployment* as its target. Increasing effective demand for consumer goods will increase employment – and consequently also decrease unemployment – without increasing inflation at all as long as production is restricted by sales in all firms producing consumer goods. On the other hand, increasing effective demand for consumer goods, for example by reducing interest rates or tax rates, will not necessarily be distributed evenly

between industries producing consumer goods. This means that production and employment will sooner or later be restricted by capacity and not sales in at least some firms, and then inflation will begin to go up for at least some consumer goods.

In the long run production and employment in the consumer industry can increase with increasing effective demand for consumer goods if firms producing consumer goods invest in increased capacity. Of course, it may take some time to install additional capacity, and firms have to be convinced that the increase in effective demand is sufficiently large and also permanent in order to make a return to non-inflationary mark-up pricing profitable. This creates a “threshold” in two dimensions – time and effective demand – in the evolution of employment and inflation. First, increasing capacity may take so long that inflation does not return to normal levels until after several years. Second, firms may require large increases in effective demand before they install additional capacity.

A crucial question is consequently how sensitive capacity is to increases of effective demand in the consumer industry. If for some reason capacity is fixed, then the only effect of increasing effective demand beyond full capacity utilization is increasing inflation. If on the other hand capacity is sensitive to increasing effective demand, so that capacity increases quickly at moderate increases of effective demand, then increasing effective demand may not increase inflation at all in the long run, even if it increases employment and reduces unemployment.

### **5.6 Unemployment due to a fall in effective demand**

An increase in unemployment is usually initiated by a fall in market employment, and a fall in market employment often reflects a “recession”, that is, a fall in effective demand. And then the question is if a government can and should counteract the fall in spending by stimulating effective demand for some time.

The answer depends on how long the decline in effective demand is expected to last, how large the decline is, what tools the government can use, how fast they can be implemented and retracted, but also, and above all, on the type of spending that is affected: exports, investment or consumption. Government spending is a special case. Of course, a sudden fall in spending is possible even in this case, for example triggered by a budget deficit or a public debt becoming too large, but government spending can also be a tool of stabilization policy.

Suppose, *first*, that a fall in effective demand starts with a fall in exports. A central bank can lower its policy rate or even start selling its currency in order to depreciate the currency and hence reduce the prices faced by foreigners in order to compensate a falling total global

demand by increasing market shares. However, such measures may, if they have effect, meet with reactions from other countries. A government can also try to intervene in wage setting to reduce labour costs and hence export prices in another attempt to increase market shares. However, this will probably not only meet with reactions from other countries but also reduce the effective demand for consumer goods.

Alternatively, a government can subsidise investment in inventories or maintenance of machines or training of employees in order to support an anticipated recovery of global demand in the near future. But a government can also accept a temporary decline in employment and only see to it that unemployment insurance (UI) prevents employees on temporary layoff to leave the export industry before demand returns. With generous UI the decline in effective demand for consumer goods will also be moderated. Thus, even if economic policy cannot do much to counteract a (temporary) decline in the demand for export goods, it can at least avoid making the situation worse and even moderate the indirect effect of a decline in the demand for export goods on the demand for consumer goods.

Suppose, *second*, that a fall in effective demand starts with a fall of investment in real capital in the market sector. Investment in an industry depends on the phase of the product cycle (innovation, expansion, maturity or decline) and a decline in total investment may consequently be the fortuitous result of related or unrelated developments in different industries. Investment can be construction of new buildings (factories, office buildings or apartment buildings) but also purchases of new machinery or spending on R&D. A fall in the effective demand for buildings from the market sector may be counteracted by a rise in the effective demand for infrastructure from the public sector, thus stabilizing production and employment in the construction industry. Fluctuations in other types of investment in the market sector may be much harder to moderate by economic policy. Note also that a fall in investment is not necessarily associated with a fall in employment. For example, if a fall in investment is dominated by a fall in imports of machines the effect on employment may be negligible.

The demand for investment in real capital depends on its expected profitability compared to the expected profitability of investing in securities. Investment also presupposes financing, and if this is a bottleneck, monetary policy can help by allowing banks to lend by creating money or, if banks find financing of investment in real capital too risky, by guaranteeing such lending. In this way monetary policy can influence the *effective* demand for investment goods in the market sector.

Economic policy can also moderate *indirect* effects of a fall in investment on employment. If, for example, the construction industry is hit by a loss of orders from the market sector, and this loss is not fully compensated by orders from the public sector, then the resulting layoffs in the construction sector may also reduce sales, production and employment in the consumer industry unless the loss of earnings is compensated by unemployment insurance.

Moreover, suppose that a fall in aggregate effective demand starts with a drop in consumption. Such a fall can be caused by a fall in “consumer confidence”, as in the U.S. in 1990-91, as discussed, for example, by Blanchard (2009 p. 77). Tax cuts or lower policy rates do not necessarily moderate the fall, since higher incomes may only lead to more saving when consumers face uncertainty.

Unemployment insurance (UI) and the tax system automatically moderate changes in effective demand for consumer goods during a business cycle, with UI adding to expenditures during a recession and taxes restricting them during a boom. The effect on consumer spending of these “automatic stabilizers” during a recession depends very much on the design and generosity of the UI-system.

A government can also add to effective demand by “discretionary” fiscal policy, that is, temporary expenditure increases or tax reductions. But such additions to the government budget usually needs time-consuming approval by the legislative assembly. It may also take some time before decisions can be implemented and have effect on effective demand. It can even be argued that discretionary fiscal policy is likely to have its biggest impact when the recession is over.

Monetary policy also has a time lag, but stabilization policy is usually identified with monetary policy, with discretionary fiscal policy kept as a reserve for deep – and long – recessions, especially when not even interest rates close to zero can stimulate investment or consumption. But what can fiscal policy do during a deep recession?

Tax reductions can stimulate consumption expenditures, but in practice the effect can be small unless tax reductions are granted to low-income households which spend the money instead of saving it. But note that during a deep recession with an increasing risk for becoming unemployed, a tax reduction may be saved instead of spent, as a precaution, even by households with low incomes, especially when unemployment benefits are small.

A more effective type of fiscal policy during a deep recession is expansion of government expenditures. A particularly effective type of fiscal policy is to prevent reductions of public

expenditures caused by “automatic stabilizers”, like enforcing balanced budgets for local governments in a recession (since local taxes automatically fall when incomes fall).

It is usually assumed that fiscal policy during a recession has to be temporary, like public works in the thirties. This can make it hard to find meaningful projects, especially since they should not “crowd out” incipient private investment. It has sometimes even been suggested that it doesn’t matter much what workers do in such projects, and that what matters instead is the preservation of professional skills apart from stabilizing incomes.

The basic reason for restricting fiscal policy to temporary measures is, of course, that the decline of effective demand from the private sector during a recession is assumed to be temporary. But is this assumption always valid? What if there can be not only temporary but also permanent declines in effective demand for consumption goods and investment goods? One possibility is increased preferences for savings in a developed economy with deficient public provisions for health care, unemployment, and old age, as discussed, for example, by Koo (2008) for Japan. Another possibility is increasing inequality, as discussed, for example, by Stiglitz (2012) for the U.S., with effective demand restricted by low and stagnating incomes for most households, and so high incomes for others that only a very small fraction can be spent on goods and services.

If government deficits have to be financed by borrowing from the private sector (banks, firms, and households), an additional restriction on fiscal policy during a recession is the size of the government debt. For, if the size of the government debt is large and rapidly growing, the interest rate as well as the burden of interest payments may soon become too large and put an end to a government’s ability to borrow from the market at reasonable interest rates.

However, a government can alternatively finance its expenditures, or more precisely its budget deficits, by borrowing from its central bank. By lending to its government, a central bank creates money in the same way as a commercial bank creates money by lending to households and firms. Moreover, while increasing private debt suddenly can turn into a debt crisis, increasing government debt to its central bank is innocuous, since it is a debt by a government to itself. This debt doesn’t have to be repaid unless the government during a boom finds it necessary to reduce inflation by reducing aggregate expenditures through a budget surplus. Of course, a government may even today be prohibited from selling bonds directly to its CB and also prohibited from borrowing directly from its CB (including overdrafts on its account at the CB). However, as emphasized, for example, by Wray (2012), these restrictions on a sovereign state are self-imposed. They may be particularly costly during a deep recession.

A recession can be especially deep and long if initiated by a financial crisis involving a boom and a crash in security markets or real estate markets. This is because the boom makes it tempting for firms and households to finance asset purchases with borrowed money – and tempting for commercial banks to finance the loans with money creation – while the crash leaves them with debt exceeding the value of their assets. This will lead to bankruptcies in some cases and otherwise to repayment of loans until balance sheets have been repaired. As emphasized by Koo (2008, 2015), monetary policy has no effect during such “balance sheet recessions”. Not even interest rates close to zero will make spending on goods financed by new borrowing attractive when firms and households focus on reducing their debts.

### **5.7 Unemployment not due to a fall in effective demand**

An increase of unemployment may be due to a fall of employment in an industry caused by a fall of effective demand for the output from this industry, as discussed above. But a fall in employment may also be due to a substitution of real capital for labour. Or an increase in unemployment may be due to an increase in labour supply. In these cases the increase in unemployment is not *initiated* by a fall in effective demand. And then the relation between (lack of) effective demand and unemployment is less obvious.

However, unemployment cannot be reduced by increased employment unless spending on output in some industries is increased (an exception to this rule is discussed in the last paragraph of this section). Thus, increased spending on goods or services is normally a necessary condition for reducing unemployment. Firms will not increase their employment unless their sales increase.

Lower labour costs can increase employment, but only if they by reducing prices also increase sales, production and employment in some firms without reducing sales in other firms. Consider, for example, the effect of lower payroll taxes for youth employed in restaurants. These may perhaps stimulate some substitution of young workers for older workers, but they will only increase total employment if they also reduce prices so much that sales increase. Note also that increased spending on restaurants may reduce sales and hence also employment in other consumer industries.

Moreover, in some industries firms may hire assistants to help with routine tasks if the pay is sufficiently low. The basic idea must be to reduce labour costs by a more elaborate division of labour. But to do this, more hours of low-skilled/low-wage jobs must reduce hours of low-skilled/normal-wage jobs, or hours of high-skilled/high-wage jobs. These are substitutions

which, of course, may increase total employment in a firm. But they will not necessarily reduce labour costs so much that they by reducing prices also increase sales.

The hiring of low-skilled workers at low wages may sometimes increase employment without increasing sales. Consider, for example, the hiring of extra personnel to help customers fill their bags in supermarkets. This is a form of marketing, that is, attempts to increase market shares at given prices. If one of the firms in a market introduces this form of service, its competitors will usually have to follow suit and, unless total sales are increased, the net result will only be that profits are reduced for all firms. Thus, this type of hiring of low-skilled workers will hardly be advocated by the industry's employers.

### 5.8 Relation to the literature

As emphasized by Devine and Kiefer (1991 p. 307), search theory implies that policy measures can affect unemployment “in the long run” or “in equilibrium” only if they affect the flows into or out of unemployment. And relating unemployment to its flow rates is certainly an illuminating decomposition of the stock of unemployment. It explains, in particular, the distribution of unemployment between flow and duration and hence also between short-term and long-term unemployment. But it does not explain the stock of unemployment, as search theory suggests.

In fact, the stock of unemployment is always determined as a residual, as people without employment looking for jobs, where employment is determined by firms and the matching process. Hence unemployment is not determined by, but is a *restriction* on, the flow rates. In other words, high unemployment is not caused by long spells of unemployment but vice versa: long durations are caused by high unemployment, even “in the long run” or “in equilibrium”.

More precisely, in equilibrium defined as a steady state, with a stable stock of unemployment  $U$ , the outflow  $\lambda U$ , where  $\lambda$  is the outflow rate, must be equal to the inflow  $I$ ,  $\lambda U = I$ . This relation is equivalent to  $U = I/\lambda$ , where  $1/\lambda$  can be interpreted as the average duration of an unemployment spell. It is tempting to conclude that the stock of unemployment at a particular point in time is *determined* by past inflows and past outflow rates.

But this conclusion also implies that employment ( $N$ ) is determined as a residual, as the difference between the labour force ( $L$ ) and unemployment,  $N = L - U$ . Assuming instead that unemployment is determined as a residual, as people in the labour force without

employment – or, more precisely, as people without employment looking for jobs – unemployment is not determined by, but is a *restriction* on, the flow rates. It follows that the average duration of unemployment as measured by  $1/\lambda$  depends on the stock of unemployment,  $1/\lambda = U/I$ , where  $U = L - N$ .

By focusing on unemployment duration, search theory focuses on the probability per period of leaving unemployment through hiring, a probability which is assumed to depend on the “matching function”. And this concept is based on the assumption that the flow of hires ( $H$ ) is proportional to the stock of job openings ( $V$ ),

$$H = qV ,$$

where  $q$  is interpreted as probability per period of filling a job opening and is assumed to depend on the tightness of the labour market as measured by the ratio of job openings to unemployment.

However, hiring is much simpler than the concept of a matching function suggests. Hires are not “produced” by job openings and unemployment, as suggested by the hiring function. Hires are produced or, more precisely, *initiated* by decisions to replace separations or change employment. Some – perhaps most – of these hires are realized more or less instantaneously (e.g., by recalls of former employees), while other recruitments take some time. However, a firm controls its employment at every point in time unless unfilled jobs arise.

Note also that the effect of matching problems on employment can be measured directly at every point in time by the rate of unfilled jobs, as elaborated in Chapter 4, without referring to shifts of the Beveridge curve or the matching function. In fact, a shift of the matching function indicating longer recruitment times cannot affect employment unless it increases the number of unfilled jobs, as argued more in detail in Farm (2017c).

Unemployment may also depend on “sticky” wages in traditional textbooks. The implicit assumption in this context is that “flexible” wages would clear the labour market in the same way as “flexible” prices do in an auction market. But labour markets are not organized as auction markets. “Sticky” wages, that is, wages which are fixed for some time in contracts between employers and employees, should not be interpreted as an “imperfection” but as part of a model of real labour markets.

## Chapter 6: Wage formation

Of course, people only care about “real” wages, i.e. the purchasing power of wages in terms of consumer goods. But people do not recalculate their nominal wages by deflating them with a consumer price index, particularly not if inflation as measured by such an index is stable and low and if important expenditures, like interest and repayment of debt, are fixed in nominal terms. Excluding hyperinflation, which enforces very extreme behaviour, like spending wages the same day they are obtained, people consequently focus on money wages, evaluating them individually and implicitly at current prices of those consumer goods that are of interest.

Moreover, nobody can *set* real wages: only money wages can be set.<sup>32</sup> Thus, even if real wages certainly matter, we need a theory for how money wages are actually set. And then not only *real* but also *relative* wages matter, since it is *relative* wages which may affect individuals’ choices between occupations, industries, and firms.

But even if all workers want to increase their relative wage, not all workers can have their relative wage increased. And then the question of “legitimate” or “fair” wage differentials arises. If wage differentials are widely known they can cause discontent or even conflict unless they are accepted as legitimate or fair, and motivated, for example, by differences in productivity, education or working conditions, in practice often according to elaborate schemes of job evaluation.

However, money wages are determined not only by relative wages but also by some money wage which other wages are related to and which consequently determines the wage level. In practice this means that wage revisions in a market economy often are initiated by a firm or an industry acting as a *wage leader*, which consequently sets a ‘norm’ for other firms or other industries. And the wage leader’s change of the wage level can be motivated by increases in productivity or changes in the wage level abroad. In practice wage setting also crucially depends on whether wage rates are set unilaterally by trade unions or employer organisations, or in negotiations between employer organisations and trade unions, or in negotiations between individual workers and employers.

We consequently have to deal with two issues, namely the determination of the wage level and the determination of wage differentials, even if we here focus on (revisions of) the wage level. I begin by discussing the influence of “market forces” on wage setting in Section 1.

Then I try to ascertain the institutional determinants of wage setting by first examining two

---

<sup>32</sup> Of course, ‘contingent wages’ may in principle be determined in wage contracts, that is, money wages contingent on a specified price index, but this would create an inflationary feedback mechanism.

polar cases of wage formation. Thus we assume on one hand, in Section 2, that wages are set unilaterally by employees or their unions, and on the other hand, in Section 3, that wages are set unilaterally by employers or their organisations. In Section 4 we shall see how these polar cases can be moderated by negotiations. And finally I discuss (briefly) wage differentials in Section 5, economic policy related to wage setting in Section 6, and relation to the literature in the last section.

### 6.1 Wages and “market forces”

The basic message of the classical idea that wages are determined by “supply and demand” in the labour market is that wages are increasing with excess demand, decreasing with excess supply and stable when demand equals supply, suggesting that wages and employment are determined simultaneously “in equilibrium”. However, in the real world employment is not determined simultaneously as wages: instead employment is determined by product demand, capacity and labour productivity at prices set by firms *conditional on wages*. Moreover, labour supply and labour demand *as functions of wages* can only be measured in markets which are organized as auction markets.

However, excess demand and excess supply *at current wages* can be measured. But note that, while excess supply is measured by unemployment in labour force surveys, excess demand is measured by vacancies defined as “unfilled jobs”, and not by vacancies as usually defined in vacancy surveys, as elaborated in Chapter 4.

Of course, excess demand and excess supply as measured by vacancies and unemployment can coexist at the aggregate level, since we can have excess demand without unemployment in some labour markets and unemployment without excess demand in other markets. But note also that excess demand as measured by unfilled jobs appears to be almost negligible when unemployment is high, as suggested by the Swedish business survey on vacancies. (For example, the rate of unfilled jobs was 0.3 per cent of employment in the private sector in Sweden in 2009, when the unemployment rate was 8.3 percent).

Now, how will “market forces” influence wages or, more precisely, how will excess supply as measured by unemployment, or excess demand, as measured by unfilled jobs, influence wage revisions? Consider first unfilled jobs. In general recruitment difficulties may raise wage offers by firms but also, or alternatively, raise wage claims by job applicants and also (after some time) wage claims by already employed. Thus, labour shortage may give rise to “wage drift” by firms or “wage pressure” by workers. In other words, unfilled jobs may raise wages either directly, when higher wages are initiated by firms, or indirectly, when higher

wages are initiated by job applicants and accepted by firms, both in the short run, for new personnel, and in the long run, for old personnel.

On the other hand, note that difficulties to recruit labour of a certain type, or labour in certain regions or industries, may be met by attempts to increase labour supply by active labour market policy instead of increasing wages – which, after all, is a very costly way of attracting new personnel. And even if unfilled jobs in an industry may give rise to wage competition between firms, they may also initiate an agreement between firms to compete with other and less expensive means. Moreover, difficulties to recruit workers with certain skills may be met by attempts to replace them with other skills or new technology.

Consider next unemployment. How will, for example, unemployment of unskilled workers affect their wages. Will high unemployment reduce not only wage offers by firms but also wage claims by workers? And will lower wages reduce unemployment by increasing employment? Of course, lower wages in an industry dominated by unskilled labour may lower the price of the industry's output so much that sales, output and employment also increase. But in a developed economy this may apply only to industries like fast-food restaurants and cleaning services. In industries like manufacturing, unskilled labour no longer constitutes a major part of the labour force and can only function as complements to skilled labour to some degree, for example by being assistants. Thus, the extent of "simple jobs" in a developed economy may be limited, which means that lower wages may have a negligible effect on the employment of unskilled labour. And then unemployment may continue to reduce wage offers, wage claims and wages until a statutory minimum wage is reached.

At the aggregate level, unemployment can be interpreted as a general indicator of "wage pressure". For example, the key idea in Layard et al. (1991) is that "wage pressure" builds up unless there is an excess supply of labour or, more precisely, an "effective" excess supply, measured by  $cu$ , where  $c$  is a measure of the "search effectiveness" of the unemployed and  $u$  is the unemployment rate. And "wage pressure" means firms bidding up wages against each other or workers pressing their wage claims, which depend on firms' chances of filling their vacancies or workers' chances of finding jobs. Emphasizing unemployment as a summary measure of "market forces" is particularly relevant when there is no precise information on excess demand, that is, when vacancy surveys don't measure unfilled jobs.

However, even if "wage pressure" and "wage competition" are suggestive concepts, obviously related to wage formation, they are not sufficiently precise. When discussing wage formation or, more precisely, wage revisions, we have to be more specific. Exactly how are wages revised and how does the rate of change of money wages depend on the institutional

details of wage formation? How will wage claims by employees and wage offers by employers in an industry depend on the prevailing regime: production restricted by capacity or production restricted by sales? And to see what wage setting involves in practice, it will be informative to first consider two polar cases, one where wages are set by workers or their unions and one where wages are set by firms or their organisations.

## **6.2 Wages set by workers or their unions**

Suppose that wages are set as prices are set in most consumer markets, where prices are set by sellers while quantities are chosen by buyers. In a labour market this would mean that employers take wages as given but on the other hand are free to choose whom to employ. This hypothetical case serves as a basis for formulating or understanding wage claims by workers or their unions even in more realistic surroundings. There is an important distinction between wage setting by already employed and wage setting by job applicants and we begin by discussing the first case.

### *Wage setting by employees*

For persons already employed wage setting means *wage revision*, but this can be designed in many different ways. Suppose that every employee has a wage contract for a year and when there is time to revise the wage an employee asks for a certain wage increase, while the employer chooses between on one hand accepting the wage claim and prolonging the contract with the new wage, and on the other hand rejecting the wage claim and dismissing the worker (and hire a new worker in the same occupation). Now, given this rule, what wage increase will the employee ask? Without knowledge of her colleagues' wages or wage offers she faces a difficult dilemma. The threat of dismissal may even force the employee to ask for a wage decrease (depending on the prospects of finding another job). Information on her colleagues' past wages may make her want to have the same wage as one of her colleagues, but without knowledge of her colleagues' asking wages she is still caught in a difficult dilemma.

However, assuming that even the rules of wage setting are up to employees, they can choose to coordinate their asking wages by establishing a trade union and ask for a common increase of the wage level, subject to the restriction that the employer is free to adjust prices, production and employment to the new wage level. (And once wages have been set in a firm, its employees won't object to prices being adjusted to wages in order to maximize profits.) Assuming that not only employers but also employees want to maximize their income, what will the wage claim be?

Now, even this maximization problem is, of course, subject to restrictions, first of all the fact that increasing wages may reduce employment by increasing prices so much that sales and production will be reduced. This will increase the risk for becoming unemployed and may consequently restrain wage claims. The risk is most obvious if wage setting is decentralized to the firm level, since then the firm facing the highest increase of the wage level will have to choose between either raising prices so much that most of its customers will turn to other firms with lower prices in the industry, or to adapt to the lowest price set by other firms. In the first case the firm's employment may fall drastically and in the second case profits may fall so much that the firm is forced to close down.

However, since employees, like all agents, avoid difficult dilemmas if they can, they will organize and form a union for the whole industry. This will eliminate restrictions on wage setting caused by the risk for an individual firm to lose sales or profits to other firms within the industry when wage setting is decentralized to firms. But higher wages may still imply higher product prices and hence also lower sales, production and employment in the industry – unless there are counteracting factors. More precisely, according to Chapter 3, and as long as production is restricted by sales and not capacity at prices chosen by firms,

$$(1) \quad p = (1+m)(1+h)(w/a) \text{ and } N = D(p)/a \text{ if } D(p) < k,$$

where  $p$  denotes the price a firm sets,  $D(p)$  the firm's sales at this price,  $k$  its capacity,  $w$  the wage level,  $h$  variable costs other than labour costs measured as a share of labour costs,  $m$  a mark-up on direct costs,  $k$  the firm's capacity,  $N$  its employment in hours and  $a$  its labour productivity.

It follows that increasing product demand or decreasing prices on other inputs than labour will create a "room" for raising the wage level without raising prices or reducing employment, other things equal (including labour productivity). And if no employee wants to risk losing her job, this room will be a binding restriction on wage claims.

However, if demand falls so that  $D(p)/a$  decreases even for an unchanged wage level, then employment cannot be kept unchanged unless also wages are reduced. This creates a dilemma for an industry's trade union. But it is a dilemma which can be resolved by relaxing the employment restriction. Thus, instead of reducing wages and keep their jobs, workers may prefer to keep wages constant or even increasing while they accept some risk of losing the job – particularly if the unemployment rate is low and there are other expanding industries with unchanged or higher wages where employees can move.

On the other hand, suppose that an industry's production is restricted by capacity and not sales, so that prices are market clearing and determined by the demand for its products, while employment is determined by the industry's capacity and labour productivity. Then a higher wage level will not reduce the industry's employment until the increase is so large that mark-up pricing with production restricted by sales is more profitable than market-clearing.

Increasing nominal wages as much as possible without risking dismissal may appear tempting for an industry's employees. To see this we note that real wages in an industry with the wage level  $w$  are determined by

$$(2) \quad w/p_c = \frac{a_c(w/w_c)}{(1+m_c)(1+h_c)},$$

where  $p_c$  denotes the price level in the consumer industry according to (1) and  $w_c$  is the wage level in the consumer industry. Thus, even if workers in the consumer industry cannot increase their real wages by increasing their nominal wages, workers in other industries can, by increasing the relative wage level  $w/w_c$ .

It will consequently be tempting for workers outside the consumer industry to increase their *relative* wages. Such increases will not affect real wages in the consumer industry as long as production is restricted by sales in the consumer industry. But increasing nominal wages outside the consumer industry imply increasing demand for consumer goods, and if the demand for consumer goods becomes so large that production is restricted by capacity and prices are market clearing, then inflation follows.

Next, consider the consumer industry, that is, firms importing, producing, or distributing consumer goods. Prices of consumer goods, as well as inflation as measured by a consumer-price index, are determined by (1). A trade union for this industry can consequently control inflation as usually measured as long as production is restricted by sales at prices set by firms. Hence the union should adjust its wage claims to its target for inflation, conditional on its predictions of labour productivity, mark-ups, and input prices in the future.

Finally, consider the export industry, that is, firms with most of their sales going to foreign buyers. Increasing wages mean increasing costs for exporters, but this does not necessarily imply decreasing employment. If an exporter is a price taker in global markets, then increasing costs imply decreasing profits, but this will not reduce sales, production and employment unless the decrease in profits is so large that the firm has to close down. If a firm is a price maker in global markets, then increasing wages will also reduce profits. Employment will not fall if the firm is producing at full capacity, but if the firm's production

is restricted by sales at the price it sets, then increasing wages will imply decreasing sales, production and employment at given effective demand unless the effect on prices is counteracted by increasing labour productivity or falling prices of raw materials or other inputs. Thus, if employment is a binding restriction on a union's wage claims in the export industry, then we can define a room for revision of the wage level determined by expected developments of labour productivity, mark-ups and input prices according to (1), and in general also expected changes of global demand.

There is, however, an important complication if a wage increase consistent with non-decreasing employment in the export sector is lower than expected price increases of consumer goods. In this case unions representing workers in exporting firms may demand compensation for expected inflation even if such demands may be expected to reduce employment in the export sector.

#### *Wage setting by job applicants*

Next, consider asking wages by job applicants, including unemployed applicants. Firms recruit new personnel to replace separations or increase employment. A firm will take current wages of its employees as a point of departure for a decision to accept or reject an offer from a job applicant. Realizing this, a job applicant may ask for information on the current wage in her occupation, either from the employer or from a trade union. Otherwise an applicant may risk submitting an asking wage which is much higher than wage offers from other applicants.

Moreover, while we assume in this section that workers are free to set wages, including the rules for wage formation, we also assume that employers are free to choose between workers applying for jobs. Thus, when the applicants for a job have different wage claims, an employer is free to choose among applicants. And the employer is not forced to choose the applicant with the lowest asking wage, and if differences in qualifications matter for the job, then the employer is free to choose the applicant with the highest qualifications, even if the applicant's asking wage is the highest.

Now, in some jobs differences in qualifications above some level do not matter, and if all job applicants are "properly qualified", the lowest wage claim will be decisive for the employer. In this situation submitting wage claims independently by job applicants would be to create a dilemma implying a strong tendency for wage dumping – contrary to the interest of all applicants. And rational agents avoid difficult dilemmas when they can, so job applicants will prefer rules in the labour market which prescribe a wage predetermined by the current wage in the firm or the industry. This means that an employer with a job opening will hire the

first qualified worker she can find at a given wage, either by recalling a former employee or hiring the first applicant who contacts her, a system which is not obviously unfair to other workers or potential applicants or, at any rate, better than the alternative.

On the other hand, in some jobs individual qualifications matter so much that an employer may prefer to hire the person with the highest qualifications even if the person's wage claim is the highest – and if wage claims are submitted independently by job seekers. In this case an employer can even interpret a low wage claim as a signal of lower qualifications, or an attempt to compensate for lower qualifications with a lower wage, provided the employer assumes that job applicants assume that the employer will weigh qualifications against wages for all applicants. Thus, in this case independent wage claims would add another element of uncertainty to wage formation, apart from the risk for wage dumping. It will consequently be rational for all applicants to ask for the same wage and let the employer's choice be based entirely on the qualifications of the applicants and not on their different wage claims.

### **6.3 Wages set by employers or their organizations**

Higher wages imply lower profits for a firm, so a firm wants its wages to be “as low as possible”. What does this imply for a firm's wage offers when wages are to be revised? This obviously depends on the alternatives available for its employees and on other firms' wage offers. If wage offers differ substantially between firms in an industry, then firms with low wage offers may soon find some of their workers leaving for employment in other firms, while firms with high wage offers may regret their offers as being unnecessarily high. And without coordination wage offers are likely to differ, particularly if they are submitted independently by firms. Since employers avoid difficult dilemmas if they can, and since we assume in this section that wage setting is entirely up to the employers, we can assume that all firms in an industry want to coordinate their wage revisions, implying a common agreement on changing the wage level. But what change will an industry's employers prefer?

This also depends on the alternatives which employees have. In an emerging economy the only alternative may be subsistence farming, which means that both wage increases and original wage levels can be moderate. In an advanced economy even an unchanged wage level can provoke some workers to leave the industry – if wage levels in other industries are revised upward. But why would other industries want to offer higher wage changes? An industry's preferences depend on the phase of its product cycle (innovation, expansion, stagnation or decline). An expanding industry wants to attract new employees quickly and this can be done by a higher wage level. If the current wage level is insufficient it may be tempting to offer a

higher wage increase in per cent than others, but it may be sufficient to do this only once, if the new (relative) wage level is sufficient to attract a persistent flow of new personnel.

Of course, different wage offers from different industries may also result if wage offers are made independently. But if there is an employer organisation for every industry, facilitating coordination among an industry's firms, we can also assume that there is at least some coordination among the employer organisations. For example, all industries may accept one industry as a wage leader. And an industry with a declining demand for its products may find it particularly easy to accept moderate wage increases and declining employment.

Now, if employers can avoid wage competition between firms in an industry and even wage competition between industries, they should be able to be very restrictive when wages are revised. But preferences differ. On one hand the consumer-goods industry will probably object to an overall reduction of wages in the economy, since this would reduce the effective demand for consumer goods. On the other hand, reducing the economy's wage level will not reduce the effective demand for export goods, while the costs of the export industry are reduced, which will increase either profits or the industry's competitiveness (i.e. making it possible to reduce prices if necessary).

It remains to discuss employers' wage offers to job applicants. An employer empowered to set wages without negotiations may be tempted to ask for independent bids. As in Section 2 this may be unproblematic if qualifications above a certain level do not matter for the job, but if they do the employer have to consider the trade-off between wage offers and qualifications. Another complication is that if this procedure succeeds in producing a lower wage than employees with the same qualifications have in the firm – which must be the purpose of the procedure – then discontent may sooner or later follow – unless the employer manages to keep wage differentials a secret to all employees. Such complications can be avoided by offering all qualified applicants the same wage as already employed persons with the same qualifications have, and then hire the first qualified person to look for the job or the most qualified person.

To sum up, employers basically want to set wage levels “as low as possible”. Wage competition between firms and between industries may raise wage levels, but if it does, it can also initiate attempts to coordinate wage setting in order to eliminate or moderate wage competition. The export industry may be particularly anxious to reduce wages, since this would not reduce the demand for its products, while on the other hand the consumer-goods industry may object to lower wage levels, particularly in other industries. Note also that wage

competition may be a very costly way of attracting labour compared to other methods, like active recruitment or on-the-job training.

#### **6.4 Wages moderated by negotiations**

A change of the wage level in an industry is not necessarily the outcome of negotiations between a trade union and an employer organisation. It may alternatively be set unilaterally by either a trade union or an employer organisation. Or it may not be set at all, meaning that it is the aggregate result of independent agreements between individual employers and individual employees without any restrictions on the resulting aggregate outcome.

Revising wages by individual negotiations is, of course, the only possibility in new industries without trade unions or employer organisations. But individual negotiations may also be the outcome of a collective agreement if both the trade union and the employer organisation find them advantageous. And the reason for accepting such an agreement must for the trade union be expectations of higher wages than otherwise (for example due to labour shortage), particularly if there is a wage leader in the economy defining a ceiling for wage increases which is unacceptable to the union, while the employer organisation may want to revise the industry's wage structure.

Now, when analysing wage revisions in a particular industry we have to distinguish between four different categories, namely: 1) an industry with both a trade union and an employer organisation; 2) an industry with a trade union but no employer organisation; 3) an industry with an employer organisation but no trade union; and 4) an industry with neither a trade union nor an employer organisation.

A fundamental aspect of the institutional structure within which wages are revised is how an economy's industries are distributed between these categories. One of these types may be dominant, for example type 1, representing a fully institutionalized labour market, or type 4, representing a completely free market. The existence of industries of type 2 or type 3 may be difficult to explain a priori – since the establishment of either a trade union or an employer organisation should provoke the establishment of a counterpart – and may be due to special circumstances. In any case, we have already discussed wage formation in industries of type 2 (in Section 2) and in industries of type 3 (in Section 3).

Industries of type 4 are particularly interesting, since the aggregate outcome of wage revisions in this case is difficult to predict a priori. On one hand employers may succeed in their attempts to keep wages “as low as possible”, but on the other hand wage competition between firms may result in excessive wage drift. This makes it important for a government at

least to *measure* changes of the wage level in industries of type 4 and intervene if either wage deflation or wage inflation in such industries threatens the economy's stability.

In industries of type 1 negotiation may result in a compromise between employee objectives as discussed in Section 2 and employer objectives as discussed in Section 3. A bid from the trade union implying, for example, an increase of the wage level by 3% may be motivated by increasing labour productivity or decreasing input prices, as elaborated in Section 2. The employer organisation may answer that these predictions are unrealistic, but that 2% is compatible with realistic predictions and hence acceptable.

A more serious conflict may arise if the trade union demands compensation for expected inflation, and inflation, as measured by a consumer price index, is expected to be 5%. The employer organisation may then answer that this would most certainly lead to lower employment, and if the union finds this prediction at least partly convincing it may reconsider its bid. Both parties may finally agree on say, 3%, each party believing that uncertainties will ultimately be resolved in its favour.

An even more serious conflict may arise in a boom if production is restricted by capacity and not sales, since in this case higher wages will only reduce profits. This applies up to a point, as discussed in Section 2, but if effective demand is so large that the wage level can increase by, for example, 6% without reducing employment, then the parties will face a zero-sum game with an uncertain outcome. Of course, employers may predict that profits will increase substantially even with a 6% increase of the wage level, but on the other hand the trade union cannot be quite certain of the outcome, so a compromise of 5% may be acceptable to both parties in the industry – even if it is not acceptable to the parties in other industries.

This brings us to the issue of coordination of wage revisions between industries. Since people avoid difficult dilemmas if they can, we can perhaps exclude the idea of industry agreements made independently without any coordination at all between industries. Let us discuss two possibilities, namely central coordination and wage leadership.

Central coordination presupposes both a powerful Federation of Trade Unions and a powerful Federation of Employer Organisations. Negotiations between such federations will not necessarily result in an increase of the wage level common to all industries. In fact, a central agreement may also include permissions or prescriptions for some industries to deviate upward or downward from the average increase of the wage level. The reason for such a deviation may, for example, be labour shortage in an industry which only can be reduced by adjustment of relative wages. Agreements on such deviations are probably difficult but not

impossible, particularly not if they are supported by arguments which are convincing to the parties in all industries.

An alternative to central coordination is coordination by wage leadership by one of the industries, for example the export industry or the consumer industry. This may arise either if central coordination cannot be established in the first place, or, when such a system exists to begin with, if one of the federations finds the system counter-productive and dissolves itself. Coordination by wage leadership means that wage revision starts in the leading industry and all other wage revisions have this industry's agreement on the increase of the wage level as a ceiling, or at least as a bench-mark from which one cannot deviate without very strong arguments which other industries can accept, in particular the parties in the wage-leading industry.

### **6.5 Wage differentials**

Even if the most basic wage differentials are between occupations, wage differentials between industries are also important, in addition to wage differentials between the market sector and the public sector. In some cases wage differentials between industries and sectors may apply to the same occupations, but some occupations may exist in only one industry or one sector. If wage differentials between occupations are to be important "signals" in the economy, guiding people in their job choices, they have to be well-known and also stable over time.

Wage differentials between occupations can seldom be explained by different productivity, since different occupations are mostly not substitutes but complements in an economy's division of labour. But different occupations are defined by different tasks, which in general require different levels or types of education, different training, and different experiences or careers or, in summary, different costs in terms of time and money which must be compensated by wage differentials in order to attract supply to those industries whose products are demanded. Differences in working conditions may sometimes also lead to compensating wage differentials.

Wage differentials between managers and workers have increased enormously in recent years for reasons that are not quite clear. Increased competition between firms for managers is one possible explanation, but this competition must also be based on a belief that a firm's future crucially depends on the ability of some individuals to find the best strategic choices in a world of uncertainty.

A new and expanding industry may introduce new occupations with new wage levels, which may or may not differ substantially from the wage levels in other comparable

occupations. The expansion of the financial industry after 1980, based on new technology and new regulations, offers particularly interesting examples. Occupations with a large number of employees may have lower wage levels than occupations with a small number of employees in capital-intensive industries where total costs are less sensitive to labour costs (even with comparable skills).

Wage differentials between the public sector and the market sector are particularly sensitive to political decisions. Public institutions may use wage levels in the market sector as bench-marks, and may even accept the market sector as a wage leader by refusing to revise wages until wages have been revised in the market sector. Wage differentials between individuals in the same firm and with the same occupation may be explained by different productivity as measured objectively by piece rates or estimated subjectively by employers and accepted by employees in individual negotiations as legitimate work incentives.

## **6.6 Economic policy**

Since changes of the wage level are crucial events in the evolution of a market economy, we are interested not only in their determinants but also in the possibility to control wage inflation by economic policy. Suppose that the target for wage inflation is 2%. Suppose also that this is a target not only for the average change of the wage level but also for wage inflation in the consumption-goods industry. Now, inflation as measured by a consumer-price index will be determined by the development of labour productivity, input prices and mark-ups in addition to wages in the consumer industry. If only wages change, then inflation will also be equal to 2%. Increasing labour productivity or decreasing input prices would reduce inflation below 2%, perhaps even below 0%, but in this case, when wages are increasing, price deflation would hardly be a problem for wage earners.

Moreover, price inflation cannot be controlled except by controlling wage inflation, so wage inflation is the basic control variable even if a target for price inflation is the ultimate goal. Now, it is usually assumed that a CB can control wage inflation indirectly by controlling effective demand and unemployment, as discussed in Chapters 3 and 5. But wage inflation can also be controlled indirectly by other means, for example by designing the legal and institutional framework so that “optimal” wage inflation is established spontaneously – or “endogenously” – by negotiations between employers and employees.

A basic assumption in this context is that both individual employers and individual employees normally prefer non-wage competition to wage competition in the labour market. Thus, both employers and employees prefer to take wages, or more precisely wage levels, as

given (even if some deviations may be established in individual negotiations). More precisely, employees prefer to compete for jobs by high qualifications instead of low wages (wage dumping), while employers prefer to attract personnel by active recruitment instead of high wages (wage competition). Note also that non-wage competition promotes the development of skills in workers and working conditions in firms.

Excluding market wages set by auctioneers or match makers, the establishment of market wages presupposes coordination, either unilaterally by trade unions or employer organisations, or in negotiations between trade unions and employer organisations. Since wage setting by trade unions may be biased upwards while wage setting by employer organisations may be biased downwards, negotiated increases of the wage level may have the highest probability of achieving a compromise of about 2%, particularly if negotiations are supported by mediators.

### **6.7 Relation to the literature**

As already noted, in the last section in Chapter 4, in traditional macroeconomics real wages are determined simultaneously as employment “in the intermediate run”, either by equality between supply and demand (in classical models) or by the intersection between the price-setting curve and the wage-setting curve (in modern models). Thus, in modern models the real wage level is determined essentially by labour productivity, since the price-setting curve is usually horizontal. This is also basically what eq. (2) above suggests, emphasizing, in addition, that it is labour productivity in the consumer industry which matters, and also the proportion of workers in other industries and their relative (nominal) wages.

## References

- Acemoglu, Daron, and James A. Robinson 2013. *Why Nations Fail: The Origins of Power, Prosperity and Poverty*, Crown Business.
- Arrow, Kenneth J., and F.H. Hahn 1971. *General Competitive Analysis*, North-Holland.
- Benassy, Jean-Pascal 1991. 'Monopolistic competition', in Hildenbrand, Werner and Hugo Sonnenschein (eds), *Handbook of Mathematical Economics, Volume IV*, 1997-2045, North-Holland.
- Blanchard, Olivier 2009. *Macroeconomics*, fifth edition, Pearson Education International.
- Blanchard, Olivier Jean, and Stanley Firscher 1989. *Lectures on Macroeconomics*, The MIT Press.
- Blinder, Alan S., Elie R. Canetti, David E. Lebow, and Jeremy B. Rudd 1998. *Asking About Prices. A New Approach to Understanding Price Stickiness*, Russel Sage Foundation.
- Boulding, Kenneth E. 1941. *Economic Analysis*, Harper and Brothers.
- Carlin, Wendy, and David Soskice 2015. *Macroeconomics: Institutions, Instability, and the Financial System*, Oxford University Press.
- Cecchetti, Stephen G. 2008. *Money, Banking, and Financial Markets*, second edition, McGraw Hill.
- Cecchetti, Stephen G. and Schoenholtz, Kermit L. 2011. *Money, Banking, and Financial Markets*, McGraw Hill.
- Devine, Theresa J., and Nicholas M. Kiefer. 1991. *Empirical Labor Economics. The Search Approach*. Oxford University Press.
- Fabiani, Sivia, Claire Loupias, Fernando Martins, and Roberto Sabbatini (eds.) 2007. *Pricing Decisions in the Euro Area. How Firms Set Prices and Why*, Oxford University Press.
- Farm, Ante 2017a. 'Pricing and price competition in consumer markets', *Journal of Economics*, 120:119-133, doi:10.1007/s00712-016-0503-7.
- Farm, Ante 2017b. 'Labour Demand and Real Wages', SOFI Working Paper 4/2017, [http://www.sofi.su.se/polopoly\\_fs/1.329901.1493105471!/menu/standard/file/SOFI%20Working%20Paper%204\\_2017.pdf](http://www.sofi.su.se/polopoly_fs/1.329901.1493105471!/menu/standard/file/SOFI%20Working%20Paper%204_2017.pdf).
- Farm, Ante 2017c. "Measuring the Effect of Matching Problems on Unemployment", manuscript available at <https://researchdatabase.su.se/converis/portal/Person/124693>.
- Farm, Ante 2017d. 'Basic Monetary Economics', SOFI Working Paper 8/2017, [http://www.sofi.su.se/polopoly\\_fs/1.361169.1513000141!/menu/standard/file/SOFI%20Working%20Paper%208\\_2017.pdf](http://www.sofi.su.se/polopoly_fs/1.361169.1513000141!/menu/standard/file/SOFI%20Working%20Paper%208_2017.pdf).

- Gordon, Robert J. 2016. *The Rise and Fall of American Growth*, Princeton University Press.
- Hall, R.L. and Hitch, C.J. 1939. "Price Theory and Business Behaviour", *Oxford Economic Papers*, 2, 12-45.
- Jones, Charles I. 2014. *Macroeconomics*, Third Edition, W. W. Norton & Company.
- Kalecki, Michal 1971. *Selected Essays on the Dynamics of the Capitalist Economy 1933-1970*, Cambridge University Press.
- Keynes, John Maynard 1936. *The General Theory of Employment Interest and Money*, Macmillan.
- Koo, Richard C. 2008. *The Holy Grail of Macroeconomics: Lessons from Japan's Great Recession*, Wiley.
- Koo, Richard C. 2015. *The Escape from Balance Sheet Recessions and the QE Trap*, Wiley.
- Krugman, Paul 2012. *End This Depression Now*, Norton.
- Lavoie, Marc 2014. *Post-Keynesian Economics. New Foundations*, Edward Elgar.
- Layard, Richard, Stephen Nickell, and Richard Jackman 1991. *Unemployment. Macroeconomic Performance and the Labour Market*, Oxford University Press.
- Marks, Robert B. 2015. *The Origins of the Modern World: A Global and Environmental Narrative from the Fifteenth to the Twenty-First Century*, Third Edition, Rowman & Littlefield.
- Mian, Atif, and Amir Sufi 2014. *House of Debt. How They (and You) Caused the Great Recession, and How We Can Prevent It from Happening Again*, The University of Chicago Press.
- Minsky, Hyman 2008. *Stabilizing an Unstable Economy*, McGraw-Hill (first printed in 1986).
- Nagle, Thomas,T. and John E. Hogan, 2006. *The Strategy and Tactics of Pricing*, Fourth Edition, Pearson Education.
- Okun, Arthur, M. 1981. *Prices and Quantities: A Macroeconomic Analysis*. Basil Blackwell.
- Piketty, Thomas 2013. *Capital in the Twenty-First Century*, translated by Arthur Goldhammer, The Belknap Press of Harvard University Press.
- Scherer, F.M. 1980. *Industrial Market Structure and Economic Performance*, Second Edition, Houghton Mifflin.
- Simon, Hermann, 1989. *Price Management*, North-Holland.
- Stigler, George J. 1947. "The Kinky Oligopoly Demand Curve and Rigid Prices", *Journal of Political Economy*, 55, 432-449.

Stiglitz, Joseph E., 2012. *The Price of Inequality. How Today's Divided Society Endangers Our Future*, Norton.

Stiglitz, Joseph E., Amartya Sen and Jean-Paul Fitoussi, 2010. *Mismeasuring our lives. Why GDP doesn't add up*. The report by the commission on the measurement of economic performance and social progress. The New Press.

Turner, Adair, 2016. *Between Debt and the Devil. Money, Credit, and Fixing Global Finance*, Princeton University Press.

Wicksell, Knut, 2010. *Lectures on Political Economy, Volume II: Money*, Routledge Revivals (first edition published in 1906 in Swedish).

Wray, L. Randall, 2012. *Modern Money Theory. A Primer on Macroeconomics for Sovereign Monetary Systems*, Palgrave Macmillan.